

OS Homo sapiens.
 XX WO200155314-A2.
 XX PD 02-AUG-2001.
 XX PF 17-JAN-2001; 2001WO-US01324.
 XX 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218250.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225457.
 PR 14-AUG-2000; 2000US-0225758.
 PR 18-AUG-2000; 2000US-0225759.
 PR 22-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226688.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 08-SEP-2000; 2000US-02311242.
 PR 08-SEP-2000; 2000US-02311243.
 PR 08-SEP-2000; 2000US-02311244.
 PR 08-SEP-2000; 2000US-02311413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 12-SEP-2000; 2000US-0232081.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 21-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 25-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 26-SEP-2000; 2000US-0235484.
 PR 26-SEP-2000; 2000US-0235484.

PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 13-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239335.
 PR 13-OCT-2000; 2000US-0239337.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 01-DEC-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0251989.
 PR 05-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0251990.
 PR 03-JAN-2001; 2001US-0259678.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX PA

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 AVFFAGR 9
 Db 1081 AVFFSGR 1087

RESULT 8

AAR96948
 ID AAR96948 standard; peptide; 13 AA.

XX AC AAR96948;
 XX AC
 DT 20-NOV-1996 (first entry)
 XX
 DE Target sequence #3 for serine protease inhibitor protein.
 XX
 KW Serine protease inhibitor; thrombin; serpin; antithrombin-III; AT-III;
 KW alpha1-antitrypsin; AAT; activated protein C; aPC; heparin; thrombosis;
 KW antithrombotic agent; disseminated intravascular coagulation; therapy;
 KW anticoagulant agent; DIC; unstable angina; myocardial infarction;
 KW thrombotic stroke; pulmonary embolism; clotting disorder;
 KW serine protease.

Synthetic.

XX WO9610638-A1.

XX 11-APR-1996.

XX 12-SEP-1995; 95WO-GB02155.

XX 03-FEB-1995; 95GB-0002138.

PR 30-SEP-1994; 94GB-0019804.

XX (PPLT-) PPL THERAPEUTICS SCOTLAND LTD.

XX Carrell R, Crowther D, Hopkins PCR, Stone S;

XX WPI; 1996-209360/21.

XX Serine protease inhibitor that specifically targets thrombin - used
 to treat or prevent clotting disorders and does not require
 activation by heparin.

XX Disclosure; Page 10; 5lpp; English.

XX This sequence represents the target sequence of the serine protease
 inhibitor (serpin) protein of the invention. This sequence is derived
 from the reactive loop of antithrombin-III (AT-III), which is a natural
 serpin. The serpin does not inhibit activated protein C (aPC), does not
 require activation by heparin, and includes a target sequence (such as
 this sequence) which interacts with the proteolytically active site of
 thrombin to inhibit its activity. The inhibitor protein can be used as
 antithrombotic, or anticoagulant agents for treating or preventing
 disseminated intravascular coagulation (DIC), unstable angina, myocardial
 infarction, thrombotic stroke, thrombosis, pulmonary embolism, or other
 clotting disorders. When this protein is labelled, it can be used in
 diagnostic assays to determine the presence of serine proteases,
 especially thrombin. The protein has the same specificity as AT-III, but
 does not need to be administered with heparin, and is approximately 1500
 times more active.

XX Sequence 13 AA;

Query Match 70.5%; Score 31; DB 17; Length 13;

Best Local Similarity 77.8%; Pred. No. 27;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 STAVFFAGR 9

Db 2 STAVVIAGR 10

RESULT 9

AAR96938

ID AAR96938 standard; peptide; 19 AA.

XX AC AAR96938;

XX AC

DT 20-NOV-1996 (first entry)

XX

DE Loop swap I mutain of alaph-antitrypsin.

XX

KW Serine protease inhibitor; thrombin; serpin; antithrombin-III; AT-III;
 KW alpha1-antitrypsin; AAT; activated protein C; aPC; heparin; thrombosis;
 KW antithrombotic agent; disseminated intravascular coagulation; therapy;
 KW anticoagulant agent; DIC; unstable angina; myocardial infarction;
 KW thrombotic stroke; pulmonary embolism; clotting disorder;
 KW serine protease.

Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 3 /note= "G3A"

FT Misc-difference 4 /note= "A4S"

FT Misc-difference 5 /note= "M5T"

FT Misc-difference 6 /note= "F6A"

FT Misc-difference 7 /note= "L7V"

FT Misc-difference 8 /note= "E8V"

FT Misc-difference 9 /note= "A9I"

FT Misc-difference 10 /note= "I10A"

FT Misc-difference 11 /note= "P11G"

FT Misc-difference 12 /note= "M12R"

FT Misc-difference 14 /note= "I14L"

FT Misc-difference 15 /note= "P15N"

XX WO9610638-A1.

XX 11-APR-1996.

XX 12-SEP-1995; 95WO-GB02155.

XX 03-FEB-1995; 95GB-0002138.

PR 30-SEP-1994; 94GB-0019804.

XX (PPLT-) PPL THERAPEUTICS SCOTLAND LTD.

XX Carrell R, Crowther D, Hopkins PCR, Stone S;

XX WPI; 1996-209360/21.

XX Serine protease inhibitor that specifically targets thrombin - used
 to treat or prevent clotting disorders and does not require
 activation by heparin.

XX Claim 9; Page 35; 5lpp; English.

XX AAR96938-R96946 represent mutains of alaph-antitrypsin (ATP). These
 sequences are used as base protein moieties in the serine protease
 inhibitor (serpin) protein of the invention. The serpin also contains a
 target sequence derived from the reactive loop of a natural serpin. The
 target sequence is preferably derived from the reactive loop of
 antithrombin-III (AT-III), and has the sequence shown in AAR96937. The

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TAVFFAGR 9
|||
Db 146 TASFFAGR 153

RESULT 4
ABG28016
ID ABG28016 standard; Protein; 529 AA.

XX AC ABG28016;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #28007.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.

XX FN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR N-PSDB; AAS92203.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

XX PS Claim 20; SEQ ID No 58375; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.

XX CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 529 AA;

Query Match 75.0%; Score 33; DB 22; Length 529;

Best Local Similarity 75.0%; Pred. No. 4e+02;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 TAVFFAGR 9
|||
Db 361 TATFFAGK 368

RESULT 5
AAM50885

ID AAM50885 standard; Protein; 35 AA.

XX AC AAM50885;

XX DT 07-MAY-2002 (first entry)

XX DE Anti-bovine prion protein peptide antibody heavy chain fragment.

XX KW Prion; prion protein; PrP; PrPC; PrPSc; antibody; Fab; mouse;

XX KW ungulate; infection; bovine spongiform encephalopathy; BSE;

XX OS Mus musculus.

XX FH Key Location/Qualifiers

FT Region 1..10

FT /label= PR3

FT /note= "C-terminal end of framework region 3"

FT Region 21..31

FT /label= PR4

FT /note= "framework region 4"

XX PN WO200210335-A2.

XX PD 07-FEB-2002.

XX PF 17-JUL-2001; 2001WO-US22648.

XX PR 27-JUL-2000; 2000US-0627218.

XX PA (REGC) UNIV CALIFORNIA

XX PI (SCRI) SCRIPPS RES INST.

XX PI Prusiner SB, Safar J, Williamson AR, Burton DR;

XX DR WPI; 2002-195957/25.

XX PT Novel antibody useful for detecting infectious prions, particularly
PT pathogenic scrapie isoform of the prion protein, PrPSc in an ungulate,
PT especially cow, binds to native ungulate prion protein C in situ -
PT Example 1; Page 41; 63pp; English.

XX CC The present sequence is that of a portion (see also AAM50881 and
CC AAM50883) of the heavy chain of a recombinant Fab antibody fragment
CC (Clone C) that binds tightly to denatured bovine PrPSc but not to
CC the native conformation of the same protein in CD10 formatted ELISA.
CC The presence of PrPSc in human or animal tissues is indicative of
CC prion infection. The Clone C Fab was generated against the 96-105
CC region of bovine prion protein (PrP). Fab C reacts exclusively
CC with the single linear epitope within residues 96-105 of bovine
CC PrP. The C recombinant antibody fragments (Fabs) were isolated
CC from mouse CDNA and cloned into a vector that expressed human-mouse
CC chimeric Fabs in Escherichia coli. The purified Fabs were labeled
CC with Europium, giving an Fab designated Eu-(Hum)Fab C. This is an
CC example of antibodies of the invention that specifically bind with
CC a high degree of binding affinity to a native ungulate PrPSc in situ
CC and/or a denatured ungulate PrPSc, but not to a native ungulate
CC PrPSc in situ. The antibodies may specifically bind to epitopes of
CC PrPSc of a specific species of animal or to 1 or more types of PrPSc
CC proteins from 1 or more species of ungulates. The antibodies are
CC useful for the detection of prion infection in ungulates,
CC particularly cattle. A fast, efficient, cost-effective assay is
CC achieved. This can be used to screen for the presence of PrPSc in
CC products such as pharmaceuticals (from natural sources), food,
CC cosmetics etc. The antibodies can be used with a compound that

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 17:22:25 ; Search time 25 Seconds
(without alignments)
57.142 Million cell updates/sec

Title: US-10-014-658-15

Perfect score: 44

Sequence: 1 STAFFFAGR 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 159726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_13Jun03.*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	44	100.0	9	21	AAV44479 Human antithrombin
2	35	79.5	354	22	ABE59799 Drosophila melanog
3	34	77.3	482	23	ABE8316 Streptomyces sp. c
4	33	75.0	529	22	ABG28016 Novel human diago
5	32	72.7	35	23	AAW50885 Anti-bovine prion
6	32	72.7	46	22	AAO11922 Human polypeptide
7	32	72.7	1172	22	ABG06613 Novel human diago
8	31	70.5	13	17	AAE96948 Target sequence #3
9	31	70.5	19	17	AAE96938 Loop swap I murein

10	31	70.5	38	9	APB80061	Sequence of reacti
11	31	70.5	48	10	APB90537	Reactive site of a
12	31	70.5	71	22	AAW92770	Human digestive sy
13	31	70.5	110	19	AAW38477	Streptococcus pneu
14	31	70.5	115	24	ABU01952	S. pneumoniae pne
15	31	70.5	122	19	AAW85965	S. pneumoniae deri
16	31	70.5	131	22	ABB68778	Drosophila melanog
17	31	70.5	148	21	AAW33913	Arabidopsis thalia
18	31	70.5	193	21	AAW33912	Arabidopsis thalia
19	31	70.5	196	21	AAW33911	Arabidopsis thalia
20	31	70.5	200	21	AAW58673	Porcine reproducti
21	31	70.5	225	21	AAW06543	Arabidopsis thalia
22	31	70.5	229	21	AAW06542	Arabidopsis thalia
23	31	70.5	307	21	AAW06541	Arabidopsis thalia
24	31	70.5	379	12	AAW12163	Human PAI-1 P3-P3,
25	31	70.5	382	12	AAW11755	Human PAI-1 active
26	31	70.5	432	19	AAW59840	Mature protein seq
27	31	70.5	464	4	AAW30445	Sequence of human
28	31	70.5	464	12	AAW10390	Antithrombin III P
29	31	70.5	464	12	AAW10385	Antithrombin III m
30	31	70.5	464	12	AAW10388	Antithrombin III m
31	31	70.5	464	12	AAW10389	Antithrombin III m
32	31	70.5	464	14	AAW42895	Human antithrombin
33	31	70.5	464	14	AAW42913	Human antithrombin
34	31	70.5	464	14	AAW42916	Human antithrombin
35	31	70.5	464	14	AAW42917	Human antithrombin
36	31	70.5	464	14	AAW42918	Human antithrombin
37	31	70.5	464	21	AAW92224	Human antithrombin
38	31	70.5	464	22	AAW37963	Human antithrombin
39	31	70.5	465	20	AAW49551	Human antithrombin
40	31	70.5	510	21	AAW31822	Arabidopsis thalia
41	31	70.5	510	22	AAW32201	Arabidopsis thalia
42	31	70.5	559	22	AAW82141	S. epidermidis ope
43	31	70.5	579	21	AAW44936	Corn sulphate perm
44	31	70.5	621	21	AAW44942	Soybean sulphate p
45	31	70.5	656	21	AAW44944	Wheat sulphate per

ALIGNMENTS

RESULT 1

AAV44479

ID AAV44479 standard; peptide; 9 AA.

XX AAV44479;

AC AAV44479;

XX 27-MAR-2000 (first entry)

DT 27-MAR-2000 (first entry)

XX Human antithrombin III variant PF (residues 385-393).

DE Human, antithrombin III; ATIII variant PF; elastase-resistant;

XX IGG activated neutrophil resistant; anti-thrombin activity; heparin;

KW anti-factor Xa activity; blood clotting disorder; sepsis; trauma; stroke;

KW thrombin activation-related pathological symptom; restenosis; thrombosis;

KW acute respiratory distress syndrome; thromboembolism; reocclusion.

OS Homo sapiens.

OS Synthetic.

XX Key

XX Location/Qualifiers

FT Misc-difference 5

FT Misc-difference 6

FT Misc-difference 6

FT Misc-difference 6

FT Misc-difference 6

FT Misc-difference 6

FT Misc-difference 6

FT Misc-difference 6

FT Misc-difference 6

FT Misc-difference 6

FT Misc-difference 6

FT Misc-difference 6

FT Misc-difference 6

FT Misc-difference 6

FT Misc-difference 6

FT Misc-difference 6

FT Misc-difference 6

FT Misc-difference 6

FT Misc-difference 6

FT Misc-difference 6

FT Misc-difference 6

FT Misc-difference 6

FT Misc-difference 6


```
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 TOTPPNG 8
Db 142 TDTTPNG 148

RESULT 13
Q8MQ9 PRELIMINARY; PRT; 696 AA.
AC Q8MQ9;
DT 01-OCT-2002 (TReMBLrel. 22, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Similar to HLA-B associated transcript 2 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymph;
RA Strausberg R.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC030127; AAH30127.1; -.
DR InterPro; IPR002965; P rich extensin.
DR PRINTS; PR01217; PRICHEXTENSN.
FT NON_TER 1
SQ SEQUENCE 696 AA; 73520 MW; 9D5783506107AB37 CRC64;

Query Match 72.0%; Score 36; DB 7; Length 696;
Best Local Similarity 66.7%; Pred. No. 1.8e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 STQTPNGR 9
Db 612 SRTPTGR 620

RESULT 14
Q9V583 PRELIMINARY; PRT; 1504 AA.
AC Q9V583;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE CG809 protein.
GN CAMTA OR CG8809.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Branton R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Fosler C., Gabriellian A.B., Garg N.S., Gelbart W.M., Glasser K.,
```

```
Glodek A., Gong F., Correll J.H., Gu Z., Guan P., Harris M.,
Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
Mout S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,
Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Man K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacle J., Paragov V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003833; AAF58934.2; -.
DR FlyBase; FB00033417; Camta.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR002909; IPT_TIG.
DR InterPro; IPR000048; IQ_Region.
DR Pfam; PF00023; ank; 2.
DR Pfam; PF00612; IQ; 3.
DR Pfam; PF01833; TIG; 1.
DR PROSITE; PSS0297; ANK_REP_REGION; 1.
DR PROSITE; PSS0096; IQ; 1.
DR ANK repeat; Repeat.
SQ SEQUENCE 1504 AA; 165672 MW; 261A4383ED3BF5D6 CRC64;

Query Match 72.0%; Score 36; DB 5; Length 1504;
Best Local Similarity 75.0%; Pred. No. 4.1e+02;
```

DE Related to ser/arg-related nuclear matrix protein.

GN 93G11.140.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
RA Nyakatura G., Mewes H.W., Mannhaupt G.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA German Neurospora genome project;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL513443; CAC28666.2; -
KW Matrix protein.
SQ SEQUENCE 976 AA; 110078 MW; 9892F4376D37E4CE CRC64;
Query Match 74.0%; Score 37; DB 3; Length 976;
Best Local Similarity 75.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 TOTPPNGR 9
Db 542 TSSPPNGR 549

RESULT 6

ID Q9ZIR1 PRELIMINARY; PRT; 2157 AA.
AC Q9ZIR1;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE BAT2
GN NFKB1L1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Rowen L., Qin S., Madan A., Abbasi N., James R., Dickhoff R.,
RA Shafer L., Ratcliffe A., Loretz C., Laskey S., Hood L.;
RT "Sequence of the mouse major histocompatibility class III region."
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF109719; AAC82480.1; -
DR MGD; MGI:1340031; NFKB1L1.
SQ SEQUENCE 2157 AA; 229072 MW; 35B21F6B7C971F40 CRC64;

Query Match 74.0%; Score 37; DB 11; Length 2157;
Best Local Similarity 66.7%; Pred. No. 3.8e-02;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 STOTPPNGR 9
Db 2073 SSRTPPSGR 2081

RESULT 7

ID Q31492 PRELIMINARY; PRT; 149 AA.
AC Q31492;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Mhc class II beta (Fragment).
OS Oncorhynchus masou (Cherry salmon) (Masu salmon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
OX NCBI_TaxID=8020;

RN SEQUENCE FROM N.A.
RP TISSUE=Blood;
RX MEDLINE=96186525; PubMed=8606054;
RA Miller K.M., Withler R.E.;
RT "Sequence analysis of a polymorphic Mhc class II gene in Pacific
RT salmon";
RL Immunogenetics 43:337-351(1996).
DR EMBL; U34697; AAB01695.1; -
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003597; IG-cl.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR000353; MHC_II_beta.
DR Pfam; PF00047; ig; 1.
DR Pfam; PF00969; MHC_II_beta; 1.
DR ProDom; PD000328; MHC_II_beta; 1.
DR SMART; SM00407; IGC1; 1.
DR PROSITE; PS0835; IG_LIKE; 1.
KW Glycoprotein; MHC II; Transmembrane.
FT NON_TER 149 149
SQ SEQUENCE 149 AA; 17121 MW; 8A2F889D3AB61953 CRC64;

Query Match 72.0%; Score 36; DB 7; Length 149;
Best Local Similarity 66.7%; Pred. No. 37;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 STOTPPNGR 9
Db 79 SSMTPPSGR 87

RESULT 8

ID O77618 PRELIMINARY; PRT; 155 AA.
AC O77618;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE ALTERNATE product P19ARF.
GN P19ARF OR ARF.
OS Monodelphis domestica (Short-tailed grey opossum).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Didelphimorphia; Didelphidae; Monodelphis.
OX NCBI_TaxID=13616;
RN [1]
RP SEQUENCE FROM N.A.
RA Sherburn T.E., Gale J.M., Ley R.D.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Chan J., Robinson E.S., Atencio J., Wang Z., Kazianis S.,
RA Della Coletta L., Nairn R.S., McCarrey J.R.;
RT "Characterization of the CDKN2A and ARF genes in UV-induced
RT melanocyte-derived skin tumors of an opossum (Monodelphis
RT domestica).";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF064809; AAC23671.1; -
DR EMBL; AF236862; AAF65223.1; -
SQ SEQUENCE 155 AA; 16467 MW; 7FFEEDB931804519 CRC64;

Query Match 72.0%; Score 36; DB 6; Length 155;
Best Local Similarity 75.0%; Pred. No. 39;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 TOTPPNGR 9
Db 130 TRTPPPNGR 137

RESULT 9

Q9DS52 PRELIMINARY; PRT; 193 AA.
ID Q9DS52

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 18:03:30 ; Search time 20.3333 Seconds
(without alignments)
114.220 Million cell updates/sec

Title: US-10-014-658-14
Perfect score: 50
Sequence: 1 STQPPNGR 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL 23.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriaph.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39	78.0	642	Q8PKZ9	Q8PKZ9 xanthomonas
2	39	78.0	642	Q8P992	Q8P992 xanthomonas
3	38	76.0	1054	Q8RU62	Q8RU62 deincoccus
4	37	74.0	950	Q823A9	Q823A9 mus musculus
5	37	74.0	976	Q9C2H8	Q9C2H8 mus musculus
6	37	74.0	2157	Q921R1	Q921R1 mus musculus
7	36	72.0	149	Q31492	Q31492 oncorhynch
8	36	72.0	155	Q77618	Q77618 monodelphis
9	36	72.0	193	Q8D5S2	Q8D5S2 mus musculus
10	36	72.0	265	Q38155	Q38155 bacterioph
11	36	72.0	250	Q8FS06	Q8FS06 anabaena sp
12	36	72.0	382	Q99362	Q99362 bacterioph
13	36	72.0	696	Q8MGQ9	Q8MGQ9 homo sapien
14	36	72.0	1504	Q9V583	Q9V583 drosophila
15	36	72.0	2157	Q9S875	Q9S875 homo sapien
16	36	72.0	2157	Q96QC6	Q96QC6 homo sapien

17	35	70.0	79	4	O60385	O60385 homo sapien
18	35	70.0	130	16	Q9K009	Q9K009 neisseria m
19	35	70.0	130	16	Q9JV22	Q9JV22 neisseria m
20	35	70.0	246	3	Q9F3R7	Q9F3R7 neurospora
21	35	70.0	260	2	Q9KIW6	Q9KIW6 thiobacillu
22	35	70.0	281	12	Q8V7H1	Q8V7H1 tt virus. o
23	35	70.0	287	16	Q8DLW7	Q8DLW7 synechococc
24	35	70.0	293	5	O17222	O17222 caenorhabdi
25	35	70.0	377	5	Q21899	Q21899 caenorhabdi
26	35	70.0	429	12	Q908Q3	Q908Q3 myxoma viru
27	35	70.0	429	12	Q9DHR4	Q9DHR4 yaba-like d
28	35	70.0	433	12	Q8JTV3	Q8JTV3 lumpy skin
29	35	70.0	433	12	Q91MV9	Q91MV9 lumpy skin
30	35	70.0	433	12	Q8V3P9	Q8V3P9 swinepox vi
31	35	70.0	442	10	Q8LRG6	Q8LRG6 oryza sativ
32	35	70.0	539	10	Q9SK50	Q9SK50 arabidopsis
33	35	70.0	590	3	Q9HE30	Q9HE30 neurospora
34	35	70.0	789	16	Q985X8	Q985X8 rhizobium 1
35	35	70.0	1073	5	Q9W1T8	Q9W1T8 drosophila
36	35	70.0	1233	16	Q9R132	Q9R132 streptomyce
37	35	70.0	1405	3	Q9P3S5	Q9P3S5 neurospora
38	35	70.0	1560	4	O60275	O60275 homo sapien
39	34	68.0	30	11	Q99KT8	Q99KT8 mus musculu
40	34	68.0	70	15	Q9YNK6	Q9YNK6 human immun
41	34	68.0	130	10	Q8S3P7	Q8S3P7 oryza sativ
42	34	68.0	142	7	Q31491	Q31491 oncorhynch
43	34	68.0	142	7	Q31489	Q31489 oncorhynch
44	34	68.0	149	7	Q31511	Q31511 oncorhynch
45	34	68.0	149	7	Q31496	Q31496 oncorhynch

ALIGNMENTS

RESULT 1

Q8PKZ9	ID	Q8PKZ9	PRELIMINARY;	PRT;	642 AA.
AC	Q8PKZ9;				
DT	01-OCT-2002	(T-EMBLrel. 22, Created)			
DT	01-OCT-2002	(T-EMBLrel. 22, Last sequence update)			
DT	01-OCT-2002	(T-EMBLrel. 22, Last annotation update)			
DE	Hypothetical protein XAC2007.				
GN	XAC2007.				
OS	Xanthomonas axonopodis (pv. citri).				
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;				
OC	Xanthomonadaceae; Xanthomonas.				
OX	NCBI_TaxID=92829;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=306 / ATCC 13902 / XV 101;				
RX	MEDLINE=22022145; PubMed=12024217;				
RA	da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,				
RA	Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,				
RA	Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,				
RA	Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,				
RA	Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,				
RA	Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.T.,				
RA	Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,				
RA	Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,				
RA	Locati E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,				
RA	Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,				
RA	Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,				
RA	Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,				
RA	Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,				
RA	Setubal J.C., Kitajima J.P.;				
RT	"Comparison of the genomes of two Xanthomonas pathogens with differing				
RT	host specificities";				
RL	Nature 417:459-463(2002).				
DR	EMBL; A0011836; AAM36869.1; -				
KW	Hypothetical protein; Complete proteome.				
SQ	SEQUENCE 642 AA; 70846 MW; 46A0656F3BB04E2B CRC64;				

DR EMBL; L33410; AA559857.1; -
 DR EMBL; U11025; AA50553.1; -
 DR EMBL; L36051; AAC37568.1; -
 DR EMBL; L36052; AAC37566.1; -
 DR EMBL; D32046; BAA06807.1; -
 DR EMBL; S76771; AAB33390.1; -
 DR EMBL; D32047; BAA21930.1; -
 DR EMBL; U59493; AAB03392.1; -
 DR EMBL; U59494; AAB03393.1; -
 DR EMBL; U59495; AAB03394.1; -
 DR EMBL; U17071; AAA74083.1; -
 DR PIR; I59281; I80105.
 DR XIM; 60044; -
 DR CO; GO:0008083; P-growth factor activity; TAS.
 DR GO; GO:0008283; P-cell proliferation; TAS.
 DR GO; GO:0007275; P-development; TAS.
 DR InterPro; IPR001323; EPO_TPO.
 DR InterPro; IPR003978; thrcmbopositin.
 DR Pfam; PF00758; EPO_TPO; 1.
 DR PRINTS; PR01485; THROMBOPTN.
 DR PROSITE; PS00817; EPO_TPO; 1.
 KW Cytokine; Glycoprotein; Hormone; Signal; Alternative splicing;
 KW Polymorphism.
 FT SIGNAL 1 21
 FT CHAIN 22 353
 FT DISULFID 28 172
 FT DISULFID 50 106
 FT CARBOHYD 197 197
 FT CARBOHYD 206 206
 FT CARBOHYD 234 234
 FT CARBOHYD 255 255
 FT CARBOHYD 340 340
 FT CARBOHYD 348 348
 FT VARSPLIC 133 136
 FT VARSPLIC 160 198
 FT VARIANT 14 14
 FT VARIANT 116 116
 FT CONFLICT 46 46
 FT CONFLICT 76 76
 FT CONFLICT 113 113
 FT CONFLICT 131 131
 FT CONFLICT 277 277
 FT CONFLICT 346 346
 SQ SEQUENCE 353 AA; 37822 MW; F0AB5449B72E5526 CRC64;
 Query Match 68.0%; Score 34; DB 1; Length 353;
 Best Local Similarity 75.0%; Pred. No. 49;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 2 TQTPNGR 9
 |||||
 DB 131 TQLPPQGR 138
 RESULT 14
 ATPG SPIOL
 ID -ATPG SPIOL STANDARD; PRT; 364 AA.
 AC P05435;
 DT 01-NOV-1988 (Rel. 05, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE ATP synthase gamma chain, chloroplast precursor (EC 3.6.3.14).
 GN ATPC.
 OS Spinacia oleracea (Spinach).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Caryophyllales; Caryophyllaceae; Chenopodiaceae; Spinacia.
 NCBI_TaxID=3562;
 OX

RN RP SEQUENCE FROM N.A.
 RC TISSUE=leaf;
 RX MEDLINE=91346685; PubMed=2151716;
 RA Mason J.G., Whitfield P.R.;
 RT "The gamma-subunit of spinach chloroplast ATP synthase: isolation and
 RL characterisation of cDNA and genomic clones."; Plant Mol. Biol. 14:1007-1018(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Monatol; TISSUE=Seedling;
 RA Oelmüller R.;
 RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 36-364 FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=cv. Nobel;
 RX MEDLINE=88211833; PubMed=2896606;
 RA Miki J., Maeda M., Mukohata Y., Futai M.;
 RT "The gamma-subunit of ATP synthase from spinach chloroplasts. Primary
 RL structure deduced from the cloned cDNA sequence."; FEBS Lett. 232:221-226(1988).
 RN [4]
 RP SEQUENCE OF 42-51; 112-128 AND 273-282.
 RX MEDLINE=96267076; PubMed=8664275;
 RA Hightower K.E., McCarty R.E.;
 RT "Proteolytic cleavage within a regulatory region of the gamma subunit
 RL of chloroplast coupling factor 1."; Biochemistry 35:4846-4851(1996).
 RN [5]
 RP DISULFIDE BOND.
 RX MEDLINE=84212589; PubMed=6233281;
 RA Moroney J.V., Fullmer C.S., McCarty R.E.;
 RT "Characterization of the cysteinyl-containing peptides of the gamma
 RL subunit of coupling factor 1."; J. Biol. Chem. 259:7281-7285(1984).
 CC -1- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
 CC GRADIENT ACROSS THE MEMBRANE. THE GAMMA CHAIN IS BELIEVED TO BE
 CC IMPORTANT IN REGULATING ATPASE ACTIVITY AND THE FLOW OF PROTONS
 CC THROUGH THE CF(0) COMPLEX.
 CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
 CC H(+) (Out).
 CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
 CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
 CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
 CC HAS THREE MAIN SUBUNITS: A, B AND C.
 CC -1- SUBCELLULAR LOCATION: Chloroplast thylakoid membrane.
 CC -1- PTM: CYS-240 AND CYS-246 ARE KNOWN TO FORM A DISULFIDE BRIDGE IN
 CC THE DARK WHICH GIVES RISE TO AN INACTIVE ENZYME; ACTIVATION CAN BE
 CC BROUGHT ABOUT BY A FERREDOXIN-DEPENDENT REDUCTION OF THE DISULFIDE
 CC BOND IN THE LIGHT.
 CC -1- SIMILARITY: Belongs to the ATPase gamma chain family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; Y00758; CAA68727.1; -
 CC EMBL; X76131; CAA53734.1; -
 CC EMBL; X17257; CAA35158.1; -
 CC PIR; S10163; PWSPG.
 CC InterPro; IPR000131; ATPase_gamma.
 CC Pfam; PF00231; ATP-synt; 1.
 CC PRINTS; PR00126; ATPASEGAMMA.
 CC TIGRFAMs; TIGR01146; ATPsyn_Figamma; 1.
 CC PROSITE; PS00153; ATPASE_GAMMA; 1.
 CC ATP synthase; Chloroplast; Thylakoid; Membrane; CF(1);
 KW Hydroxylase; Hydrogen ion transport; Transit peptide.
 FT TRANSIT 1 41
 FT CHAIN 42 364
 FT ATP SYNTHASE GAMMA CHAIN.

13-AUG-1987 (Rel. 05, Last sequence update)
15-SEP-2003 (Rel. 42, Last annotation update)
Proline-rich protein MP-2 precursor.
PRH1 OR PRP.
Mus musculus (Mouse).
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
MEDLINE=86059475; PubMed=2999141;
Ann D.K., Carlson D.M.,
"The structure and organization of a proline-rich protein gene of a
mouse multigene family.",
J. Biol. Chem. 260:15863-15872 (1985).

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; M12039; AAA40004.1; -;
DR MGD; MGI:97773; Pth1.
KW Repeat; Signal.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 261 PROLINE-RICH PROTEIN MP-2.
SQ SEQUENCE 261 AA; 26034 MW; 36E13BA7387F47D4 CRC64;

Query Match 68.0%; Score 34; DB 1; Length 261;
Best Local Similarity 66.7%; Pred. No. 35;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 STQTPNGR 9
DB 248 NTQGPQGR 256

RESULT 11
TPO_RAT
ID TPO_RAT STANDARD; PRT; 326 AA.
AC P49745;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Thrombopoietin precursor.
GN THPO
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
[1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=95331639; PubMed=7607561;
RA Ogami K., Shimada Y., Sohma Y., Akahori H., Kato T., Kawamura K.,
Miyazaki H.;
RA "The sequence of a rat cDNA encoding thrombopoietin."
RA Gene 158:309-310 (1995).
RT
CC -1- FUNCTION: LINEAGE-SPECIFIC CYTOKINE AFFECTING THE PROLIFERATION
AND MATURATION OF MEGAKARYOCYTES FROM THEIR COMMITTED PROGENITOR
CELLS. IT ACTS AT A LATE STAGE OF MEGAKARYOCYTE DEVELOPMENT. IT
MAY BE THE MAJOR PHYSIOLOGICAL REGULATOR OF CIRCULATING PLATELETS.
CC -1- SUBCELLULAR LOCATION: secreted.
CC -1- DOMAIN: TWO-DOMAIN STRUCTURE WITH AN ERYTHROPOIETIN-LIKE N-
TERMINAL AND A SER/PRO/THR-RICH C-TERMINAL.
CC -1- SIMILARITY: BELONGS TO THE EPO / TPO FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; D32207; BAA06906.1; -;
DR PIR; JC4125; JC4125.
DR InterPro; IPR001323; EPO_TPO.
DR InterPro; IPR003978; thrombopoietin.
DR Pfam; PF00758; EPO_TPO; 1.
DR PRINTS; PR01485; THROMBOPTN.
DR PROSITE; PS00817; EPO_TPO; 1.
KW Cytokine; Glycoprotein; Hormone; Signal.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 326 THROMBOPOIETIN.
FT DISULFID 50 172 POTENTIAL.
FT CARBOHYD 197 197 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 206 206 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 235 235 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 249 249 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 256 256 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 326 AA; 34556 MW; F99D7F77F896PA2C CRC64;

Query Match 68.0%; Score 34; DB 1; Length 326;
Best Local Similarity 75.0%; Pred. No. 45;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 TOTPPNGR 9
DB 131 TOLPPQGR 138

RESULT 12
TPO_CANFA
ID TPO_CANFA STANDARD; PRT; 352 AA.
AC P42705; 1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Thrombopoietin precursor (Megakaryocyte colony stimulating factor)
(C-MPL ligand) (ML) (Megakaryocyte growth and development factor)
(MGDF).
GN THPO OR TPO.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
[1]
RN SEQUENCE FROM N.A., AND SEQUENCE OF 24-44.
RP
RC TISSUE=Kidney;
RX MEDLINE=94291201; PubMed=8020099;
RA Bartley T.D., Bogenberger J., Hunt P., Li Y.-S., Lu H.S., Martin F.,
Chang M.-S., Samal B.B., Nichol J.L., Swift S., Johnson M.J.,
Hsu R.-Y., Parker V.P., Suggs S., Skrine J.D., Merewether L.A.,
Clogson C., Hsu E., Hokom M.M., Horinko A., Choi E., Pangellian M.,
Sun Y., Mar V., McNich J., Simonet L., Jacobsen F., Xie C.,
Shutter J., Chure H., Basu R., Selander L., Trollinger D., Sieu L.,
Padilla D., Trail G., Elliott G., Izumi R., Covey T., Crouse J.,
Garcia A., Xu W., del Castillo J., Biron J., Cole S., Hu M.C.-T.,
Pacifica R., Ponting I., Saris C., Wen D., Yung Y.P., Lin H.,
Bosselman R.A.;
RT "Identification and cloning of a megakaryocyte growth and development
factor that is a ligand for the cytokine receptor Mpl."
RL Cell 77:1117-1117 (1994).
CC -1- FUNCTION: LINEAGE-SPECIFIC CYTOKINE AFFECTING THE PROLIFERATION
AND MATURATION OF MEGAKARYOCYTES FROM THEIR COMMITTED PROGENITOR
CELLS. IT ACTS AT A LATE STAGE OF MEGAKARYOCYTE DEVELOPMENT. IT
MAY BE THE MAJOR PHYSIOLOGICAL REGULATOR OF CIRCULATING PLATELETS.
CC -1- SUBCELLULAR LOCATION: secreted.
CC -1- DOMAIN: TWO-DOMAIN STRUCTURE WITH AN ERYTHROPOIETIN-LIKE N-
TERMINAL AND A SER/PRO/THR-RICH C-TERMINAL.
CC -1- SIMILARITY: BELONGS TO THE EPO / TPO FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its

FT BINDING 670 670 A0 CHLOROPHYLL (BY SIMILARITY).
FT BINDING 671 671 A1 PHYLLOQUINONE (BY SIMILARITY).
FT BINDING 700 700 A1 PHYLLOQUINONE (BY SIMILARITY).
SQ SEQUENCE 734 AA; 82415 MW; 65003560BCCD986 CRC64;

Query Match 70.0%; Score 35; DB 1; Length 734;
Best Local Similarity 75.0%; Pred. NO. 70;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TOTPPNGR 9
DB 307 THTPGGR 314
|||||

RESULT 7
DIVL CAUCR STANDARD; PRT; 769 AA.
AC Q9ROQ9; Q9A292;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Sensor protein divL (EC 2.7.3.-).
GN DIVL OR CC3484.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;
OC Caulobacteraceae; Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A., AND AUTOPHOSPHORYLATION SITE.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Wu J., Ohta N., Zhao J.L., Newton A.,
RT "A novel bacterial tyrosine kinase essential for cell division and
RT differentiation";
RL Proc. Natl. Acad. Sci. U.S.A. 96:13068-13073(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldlyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA Deboy R.T., Dodson R.J., Durkin A.S., Gwinn M.B., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Utkerback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Saizberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
CC -!- FUNCTION: Required for cell division and growth. It catalyzes the
CC phosphorylation of CtrA and activates transcription in vitro of
CC the cell cycle-regulated flif promoter.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- PTM: Autophosphorylated.
CC -!- SIMILARITY: Contains 1 histidine kinase domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF083422; AAF08344.2; -;
DR EMBL; AE006007; AAK25446.1; -;
DR PIR; B87691; B87681.
DR TIGR; CC3484; -;
DR PhosSite; Q9ROQ9; -;
DR InterPro; IPR003594; ATPbind_ATPase.
DR InterPro; IPR004358; Bact_sens_pr_C.
DR InterPro; IPR003661; His_kinA.
DR InterPro; IPR005467; His_kinase.
DR Pfam; PF02518; HATPase_c; 1.

DR PFam; PF00512; HSKA; 1.
DR PRINTS; PRO0344; BCTRLSENSOR.
DR SMART; SM00387; HATPase_c; 1.
DR SMART; SM00388; HSKA; 1.
DR PROSITE; PS50109; HIS_KIN; 1.
KW Sensory transduction; Transferase; Kinase; Transmembrane;
KW Phosphorylation; Complete proteome.
FT TRANSMEM 6 26 POTENTIAL.
FT DOMAIN 547 758 HISTIDINE KINASE.
FT DOMAIN 9 221 ALA-RICH.
FT MOD RES 550 550 PHOSPHORYLATION (AUTO-).
FT CONFLICT 200 200 Q -> H (IN REF. 1).
FT CONFLICT 216 216 V -> E (IN REF. 1).
SQ SEQUENCE 769 AA; 82796 MW; 002B2428F18A57EF CRC64;

Query Match 70.0%; Score 35; DB 1; Length 769;
Best Local Similarity 85.7%; Pred. NO. 74;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 QTPPNGR 9
DB 671 QTPPGR 677
|||||

RESULT 8
RNTL MOUSE STANDARD; PRT; 1113 AA.
AC Q9EFU0; Q99PR4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Regulator of nonsense transcripts 1 (Nonsense mRNA reducing factor 1)
DE (NORF1) (Up-frameshift suppressor 1 homolog).
GN RNT1 OR HUPF1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RX MEDLINE=21067878; PubMed=11152657;
RA Medghalchi S.M., Frischmeyer P.A., Mendell J.T., Kelly A.G.,
RA Lawler A.M., Dietz H.C.;
RT "Rnt1, a trans-effector of nonsense-mediated mRNA decay, is essential
RT for mammalian embryonic viability";
RL Hum. Mol. Genet. 10:99-105(2001).
RN [2]
RP SEQUENCE FROM N.A.
RA Seig M., Strande J., Beck-Engesser G.B.J., Liehr T., Winkler T.,
RA Jack H.-M.;
RT "Genomic structure, chromosomal localization and expression of murine
RT nonsense mRNA reducing factor 1 (NORF1).";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 73-1102 FROM N.A.
RC STRAIN=129/SV;
RX MEDLINE=97008109; PubMed=8855285;
RA Perlick H.A., Medghalchi S.M., Spencer F.A., Kendzior R.J. Jr.,
RA Dietz H.C.;
RT "Mammalian orthologues of a yeast regulator of nonsense transcript
RT stability";
RL Proc. Natl. Acad. Sci. U.S.A. 93:10928-10932(1996).
CC -!- FUNCTION: Eliminates the production of nonsense-containing RNAs.
CC Essential for embryonic viability.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE DNa2/NAM7 HELICASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; X95649; CAC19480.1; --
 DR EMBL; AP005280; BAB99363.1; --
 DR InterPro; IPR001279; Blactnase-like.
 DR InterPro; IPR001587; UPF0036.
 DR Pfam; PF00753; lactamase B; 1.
 DR PROSITE; PS01292; UPF0036; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 718 AA; 77602 MW; 36D6687387C486DF CRC64;

Query Match 72.0%; Score 36; DB 1; Length 719;
 Best Local Similarity 85.7%; Pred. No. 45;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 QTPNGR 9
 Db 314 QTPDGR 320

RESULT 4

PSAB CHLVU
 ID PSAB CHLVU STANDARD; PRT; 734 AA.
 AC P56342;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Photosystem I P700 chlorophyll A apoprotein A2 (PsaB) (PSI-B).
 GN PSAB.
 OS Chlorella vulgaris.
 OG Chloroplast.
 OC Eukaryota; Viridiplantae; Chlorophyta; Trebouxiophyceae; Chlorellales;
 OC Chlorrellaceae; Chlorella.
 OX NCBI_TaxID=3077;
 RN [1]
 SEQUENCE FROM N.A.

RC STRAIN=IAM C-27 / Tamiya;
 RX MEDLINE=97303241; PubMed=9159184;
 RA Wakaugi T., Negai T., Kapoor M., Sugita M., Ito M., Ito S.,
 RA Tsudzuki J., Nakashima K., Tsudzuki T., Suzuki Y., Hamada A., Ohta T.,
 RA Inamura A., Yoshinaga K., Sugiura M.;
 RT "Complete nucleotide sequence of the chloroplast genome from the
 RT green alga Chlorella vulgaris: the existence of genes possibly
 RT involved in chloroplast division.";
 RL Proc. Natl. Acad. Sci. U.S.A. 94:5967-5972 (1997).
 CC -!- FUNCTION: PsaA and PsaB bind P700, the primary electron donor of
 CC photosystem I (PSI), as well as the electron acceptors A0, A1, and
 CC FX. PSI functions as a plastocyanin/cytochrome c6-ferredoxin
 CC oxidoreductase.

CC -!- COFACTOR: P700 is a chlorophyll A dimer, A0 is chlorophyll A, A1
 CC is a phytylquinone and FX is a 4Fe-4S iron-sulfur center.
 CC -!- SUBUNIT: A psaA/B heterodimer binds the P700 chlorophyll special
 CC pair and subsequent electron acceptors. The PSI reaction center of
 CC higher plants and algae is composed of one at least 11 subunits.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Chloroplast
 CC thylakoid membrane.

CC -!- SIMILARITY: BELONGS TO THE PSAA/PSAB FAMILY.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AB001684; BAB57928.1; --
 DR PIR; T07280; T07280.
 DR HSP; P25897; I0B0.

DR HAMAP; MF_00482; --; 1.
 DR InterPro; IPR006244; PsaB.
 DR InterPro; IPR001280; PSI_PsaA/B.
 DR Pfam; PF00223; psaa_psaB; 1.
 DR PRINTS; PRO0257; PHOTOSYPSAAB.
 DR TIGR; TIGR01336; psab; 1.
 DR PROSITE; PS00419; PHOTOSYSTEM_I_PSAAB; 1.
 KW Photosynthesis; Photosystem I; Transport; Electron transport;
 KW Chloroplast; Thylakoid; Transmembrane; Metal-binding; Iron;
 KW Iron-sulfur; 4Fe-4S; Chlorophyll.
 FT TRANSMEM 46 69 I (POTENTIAL).
 FT TRANSMEM 135 158 II (POTENTIAL).
 FT TRANSMEM 175 199 III (POTENTIAL).
 FT TRANSMEM 273 291 IV (POTENTIAL).
 FT TRANSMEM 330 353 V (POTENTIAL).
 FT TRANSMEM 369 395 VI (POTENTIAL).
 FT TRANSMEM 419 439 VII (POTENTIAL).
 FT TRANSMEM 517 535 VIII (POTENTIAL).
 FT TRANSMEM 575 596 IX (POTENTIAL).
 FT TRANSMEM 643 665 X (POTENTIAL).
 FT TRANSMEM 707 727 XI (POTENTIAL).
 FT METAL 559 559 IRON-SULFUR FX (4FE-4S) (BY SIMILARITY).
 FT METAL 568 568 IRON-SULFUR FX (4FE-4S) (BY SIMILARITY).
 FT BINDING 654 654 P700 SPECIAL PAIR CHLOROPHYLL AXIAL
 FT BINDING 662 662 LIGAND (BY SIMILARITY).
 FT BINDING 662 662 A0 CHLOROPHYLL (BY SIMILARITY).
 FT BINDING 670 670 A0 CHLOROPHYLL (BY SIMILARITY).
 FT BINDING 671 671 A1 PHYLLOQUINONE (BY SIMILARITY).
 FT BINDING 700 700 A1 PHYLLOQUINONE (BY SIMILARITY).
 SQ SEQUENCE 734 AA; 81804 MW; 7E43EDA8DEE4738 CRC64;

Query Match 72.0%; Score 36; DB 1; Length 734;
 Best Local Similarity 85.7%; Pred. No. 46;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 QTPNGR 9
 Db 308 QTPDGR 314

RESULT 5

BAT2 HUMAN
 ID BAT2 HUMAN STANDARD; PRT; 2142 AA.
 AC P48634;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Large proline-rich protein BAT2 (HLA-B-associated transcript 2) (G2).
 GN BAT2 OR G2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 SEQUENCE FROM N.A.
 RC TISSUE=T-cell;
 RX MEDLINE=90192810; PubMed=2156268;
 RA Bauerli J., Sands J., Strominger J.L., Spies T.;
 RT "A gene pair from the human major histocompatibility complex encodes
 RT large proline-rich proteins with multiple repeated motifs and a
 RT single ubiquitin-like domain.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:2374-2378 (1990).
 RN [2]
 SEQUENCE OF 1-1860 FROM N.A.
 RP MEDLINE=93272029; PubMed=8499947;
 RA Iris F.J.M., Bougueret L., Prieur S., Caterina D., Primas G.,
 RA Perrot V., Jurka J., Rodriguez-Tome P., Clavierie J.-M., Dausset J.,
 RA Cohen D.;
 RT "Dense Alu clustering and a potential new member of the NF kappa B
 RT family within a 90 kilobase HLA class III segment.";
 RL Nat. Genet. 3:137-145 (1993).
 CC -!- FUNCTION: UNKNOWN.
 CC -!- TISSUE SPECIFICITY: LIMITED TO CELL-LINES OF LEUKEMIC ORIGIN.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 16:54:27 ; Search time 4.66667 Seconds
(without alignments)
90.694 Million cell updates/sec

Title: US-10-014-658-14

Perfect score: 50

Sequence: 1 STQTPPNR 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36	72.0	351	1	ORSP_CHICK
2	36	72.0	373	1	RFL_AERPE
3	36	72.0	718	1	YU70_CORGL
4	36	72.0	734	1	PSAB_CHLVU
5	36	72.0	2142	1	BAT2_HUMAN
6	35	70.0	734	1	PSAB_WARPO
7	35	70.0	769	1	DIYL_CAUCR
8	35	70.0	1113	1	RNFL_MOUSE
9	35	70.0	1129	1	RNFL_HUMAN
10	34	68.0	261	1	PRP2_MOUSE
11	34	68.0	326	1	TPO_RAT
12	34	68.0	352	1	TPO_CANFA
13	34	68.0	353	1	TPO_HUMAN
14	34	68.0	364	1	ATPG_SPIOL
15	34	68.0	431	1	PKNA_MYCTU
16	33	66.0	204	1	YAGU_SCOLI
17	33	66.0	262	1	TRYT_DROBR
18	33	66.0	262	1	TRYT_DROME
19	33	66.0	349	1	CRMB_VARV
20	33	66.0	518	1	ATP1_HUMAN
21	33	66.0	550	1	SQAL_HUMAN
22	33	66.0	610	1	PPS_URECA
23	33	66.0	748	1	CLPE_LACLA
24	33	66.0	769	1	ZBP2_CHICK
25	33	66.0	906	1	I4G2_MOUSE
26	33	66.0	907	1	I4G2_HUMAN
27	33	66.0	907	1	I4G2_RABIT
28	33	66.0	1010	1	Y226_HUMAN
29	33	66.0	1038	1	KP17_MOUSE
30	33	66.0	4684	1	PLA1_HUMAN
31	32	64.0	258	1	PRIA_LENED
32	32	64.0	300	1	DAPA_HELPJ
33	32	64.0	311	1	NADA_SULSO

RESULT 1

ORSP_CHICK

ID ORSP_CHICK STANDARD; PRT; 351 AA.

AC P51475; P79794; 32 64.0 321 1 NADA_SULTO
 DT 01-OCT-1996 (Rel. 34, Created) 32 64.0 382 1 VE2_HPV61
 DT 01-OCT-1996 (Rel. 34, Last sequence update) 32 64.0 454 1 ERF1_DROME
 DT 16-OCT-2001 (Rel. 40, Last annotation update) 32 64.0 519 1 GLG2_SOLTU
 DE Pinopsin (Pineal opsin) (P-opsin) (Pineal gland-specific opsin). 32 64.0 580 1 SVQ_FALSO
 OS Gallus gallus (Chicken) 32 64.0 638 1 OAR1_LYMST
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; 32 64.0 647 1 GAG_SFVI
 CC Archosauria; Aves; Neognathae; Galliformes; Phasianinae; 32 64.0 962 1 LI36_CAEEL
 CC Gallus 32 64.0 1048 1 PI00_HCMVA
 CC 32 64.0 1057 1 POL_SIVAI
 CC 32 64.0 1162 1 E2K3_DROME
 CC 32 64.0 1379 1 YFF9_SCHPO

ALIGNMENTS

SEQUENCE FROM N.A.
 TISSUE=Pineal gland;
 MEDLINE=95184012; PubMed=7878470;
 Max M., McKinnon P.J., Seidenman K.J., Barrett R.K., Applebury M.L.,
 Takahashi J.S., Margolske R.F.;
 "Pineal opsin: a nonvisual opsin expressed in chick pineal.";
 Science 267:1502-1506(1995).
 [2]
 SEQUENCE FROM N.A.
 TISSUE=Pineal gland;
 MEDLINE=95059405; PubMed=7969427;
 Okano T., Yoshizawa I., Fukada Y.;
 "Pinopsin is a chicken pineal photoreceptive molecule.";
 Nature 372:94-97(1994).
 [3]
 SEQUENCE FROM N.A.
 STRAIN=White leghorn; TISSUE=Pineal gland;
 MEDLINE=98434538; PubMed=9756926;
 Max M., Surya A., Takahashi J.S., Margolske R.F., Knox B.E.;
 "Light-dependent activation of rod transducin by pineal opsin.";
 J. Biol. Chem. 273:26820-26826(1998).
 CC -!- FUNCTION: PRODUCES A SLOW AND PROLONGED PHOTOTRANSDUCTION
 RESPONSE CONSISTENT WITH THE NONVISUAL FUNCTION OF PINEAL
 PHOTORECEPTION.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- TISSUE SPECIFICITY: PINEAL GLAND.
 CC -!- PTM: SOME OR ALL OF THE CARBOXYL-TERMINAL SER OR THR RESIDUES MAY
 BE PHOSPHORYLATED (BY SIMILARITY).
 CC -!- MISCELLANEOUS: THIS OPSIN HAS APPROXIMATELY AN ABSORPTION MAXIMUM
 AT 470 NM.
 CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC OPSIN SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U15762; AAA64223.1; -

PI Guigueno A;
XX WPI; 1999-181045/15.
DR N-PSDB; AAX34085.
XX
PT Mycobacterial DNA vectors containing reporter constructs - for
PT identifying coding or promoter sequences involved in
PT infection-associated protein expression
XX
PS Claim 32; Fig 15C; 309pp; French.
XX
CC Sequences AAU04742-Y05000 and AAU07201-Y07204 represent secreted
CC proteins from various Mycobacterium species microorganisms. The
CC encoding nucleotide sequences can be used as primers and probes for
CC methods for detecting and identifying mycobacteria, especially belonging
CC to the M. tuberculosis complex. The encoded proteins can be used in
CC vaccines for immunisation against a bacterial or viral infection.
XX
SQ Sequence 99 AA;
XX
Query Match 72.0%; Score 36; DB 20; Length 99;
Best Local Similarity 75.0%; Pred. No. 1e+02; Length 99;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
XX
QY 2 TQTPNGR 9
DB 13 TRTPGGR 20
XX
RESULT 4
AAU71894
ID AAU71894 standard; Protein; 322 AA.
XX
AC AAU71894;
XX
DT 26-FEB-2002 (first entry)
XX
DE C. glutamicum metabolic pathway protein encoded by gene #29.
XX
KW Metabolic pathway protein; MP; lysine biosynthesis pathway;
KW methionine biosynthesis pathway; large-scale production of fine chemical;
KW Corynebacterium diphtheriae; diphtheria.
XX
OS Corynebacterium glutamicum.
XX
PN WO200166573-A2.
XX
PD 13-SEP-2001.
XX
PF 22-DEC-2000; 2000WO-IB02035.
XX
XX 09-MAR-2000; 2000US-187970P.
PR 23-JUN-2000; 2000US-0606740.
XX
PA (BADI) BASF AG.
XX
PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G, Kim J;
PI Lee H, Hwang B;
XX
DR WPI; 2001-582269/65.
DR N-PSDB; AAS96104.
XX
XX Nucleic acids encoding metabolic pathway proteins from Corynebacterium
PT glutamicum, useful for producing methionine and lysine in
PT Corynebacterium and Brevibacterium -
XX
PS Disclosure; Page 230-231; 316pp; English.
XX
CC The present invention relates to the isolation of novel Corynebacterium
CC glutamicum genes (AAS96073-AAS96132) encoding metabolic pathway (MP)
CC proteins. The metabolic pathway proteins of the invention
CC include enzymes involved in the lysine and methionine biosynthetic
CC pathways. The polynucleotide sequences of the invention can be used

CC for the large-scale production and/or modulation of expression of
CC fine chemicals such as lysine and methionine. The sequences of the
CC invention may be used to identify C. glutamicum and related organisms
CC e.g. C. diphtheriae in a subject to detect diphtheria.
CC AAU71863-AAU71922 represent the novel C. glutamicum metabolic pathway
CC proteins of the invention.
XX
SQ Sequence 322 AA;
XX
Query Match 72.0%; Score 36; DB 22; Length 322;
Best Local Similarity 85.7%; Pred. No. 3.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 3 QTPNGR 9
DB 314 QTPNGR 320
XX
RESULT 5
AAB79131
ID AAB79131 standard; Protein; 322 AA.
XX
AC AAB79131;
XX
DT 30-APR-2001 (first entry)
XX
DE Corynebacterium glutamicum HA protein sequence SEQ ID NO:218.
XX
KW Corynebacterium glutamicum; homeostasis; adaptation; HA protein;
KW fine chemical production; organic acid; proteinogenic amino acid;
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;
KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
KW carbohydrate; aromatic compound; vitamin; cofactor; polyketide; enzyme;
KW diagnosis; Corynebacterium diphtheriae; genetic engineering;
KW Brevibacterium; environmental condition.
XX
OS Corynebacterium glutamicum.
XX
PN WO200100842-A2.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-IB00911.
XX
XX 25-JUN-1999; 99US-0141031.
PR 08-JUL-1999; 99DE-1031636.
PR 09-JUL-1999; 99DE-1032125.
PR 09-JUL-1999; 99DE-1032126.
PR 09-JUL-1999; 99DE-1032127.
PR 09-JUL-1999; 99DE-1032128.
PR 09-JUL-1999; 99DE-1032129.
PR 09-JUL-1999; 99DE-1032226.
PR 09-JUL-1999; 99DE-1032920.
PR 14-JUL-1999; 99DE-1032922.
PR 14-JUL-1999; 99DE-1032924.
PR 14-JUL-1999; 99DE-1032928.
PR 14-JUL-1999; 99DE-1032930.
PR 14-JUL-1999; 99DE-1032933.
PR 14-JUL-1999; 99DE-1032935.
PR 14-JUL-1999; 99DE-1032973.
PR 14-JUL-1999; 99DE-1033002.
PR 14-JUL-1999; 99DE-1033003.
PR 14-JUL-1999; 99DE-1033005.
PR 14-JUL-1999; 99DE-1033006.
PR 31-AUG-1999; 99DE-1041378.
PR 31-AUG-1999; 99DE-1041379.
PR 31-AUG-1999; 99DE-1041390.
PR 31-AUG-1999; 99DE-1041391.
PR 03-SEP-1999; 99DE-1042088.
XX
PA (BADI) BASF AG.
XX
PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 17:22:25 ; Search time 25 Seconds
(without alignments)
57.142 Million cell updates/sec

Title: US-10-014-658-14

Perfect score: 50

Sequence: 1 STQPPNGR 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 159726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_13Jun03.*

1:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24:	/SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	50	100.0	9	21	AAV44478
2	27	74.0	224	22	AG01822
3	36	72.0	99	20	AAV04833
4	36	72.0	322	22	AAU71894
5	36	72.0	322	22	AAU71931
6	36	72.0	322	22	AAV79666
7	36	72.0	656	21	AAV44944
8	36	72.0	718	22	AAV91307
9	36	72.0	978	24	ABP56416

10	36	72.0	1480	22	ABBS9227
11	36	72.0	2153	22	AAU33195
12	35	70.0	57	18	AAW36504
13	35	70.0	203	20	AAV04837
14	35	70.0	345	21	AAV07022
15	35	70.0	345	21	AAV37538
16	35	70.0	381	21	AAV24617
17	35	70.0	393	21	AAV24616
18	35	70.0	395	21	AAV07021
19	35	70.0	506	20	AAV04841
20	35	70.0	539	21	AAV37537
21	35	70.0	551	21	AAV37536
22	35	70.0	905	22	ABBS9283
23	35	70.0	1118	24	ABV73900
24	35	70.0	1140	19	AAV21377
25	34	68.0	13	24	ABP81099
26	34	68.0	13	24	ABP81100
27	34	68.0	13	24	ABP81101
28	34	68.0	19	16	AAV81396
29	34	68.0	22	16	AAV76183
30	34	68.0	22	16	AAV76162
31	34	68.0	39	22	AAO12464
32	34	68.0	51	22	AAU65311
33	34	68.0	55	22	AAO12233
34	34	68.0	145	17	AAV97762
35	34	68.0	150	16	AAV80822
36	34	68.0	151	16	AAV81344
37	34	68.0	153	17	AAV01761
38	34	68.0	153	18	AAV17185
39	34	68.0	153	18	AAV17187
40	34	68.0	153	18	AAV15019
41	34	68.0	153	19	AAV77906
42	34	68.0	153	19	AAV77904
43	34	68.0	155	16	AAV79341
44	34	68.0	155	18	AAV17099
45	34	68.0	155	18	AAV15017

ALIGNMENTS

RESULT 1

AAV44478

ID AAV44478 standard; peptide; 9 AA.

AC AAV44478;

XX 27-MAR-2000 (first entry)

DT Human antithrombin III variant F2A' (residues 385-393).

XX Human; antithrombin III; ATIII variant F2A'; elastase-resistant;

XX IGG activated neutrophil resistant; anti-thrombin activity; heparin;

XX anti-factor Xa activity; blood clotting disorder; sepsis; trauma; stroke;

XX thrombin activation-related pathological symptom; restenosis; thrombosis;

XX acute respiratory distress syndrome; thromboembolism; reocclusion.

XX Homo sapiens.

OS Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 3 /note= "ATIII.NI35A Ala at 387 is substituted by Gln"

FT Misc-difference 4 /note= "ATIII.NI35A Val at 388 is substituted by Thr"

FT Misc-difference 5 /note= "ATIII.NI35A Val at 389 is substituted by Pro"

FT Misc-difference 6 /note= "ATIII.NI35A Ile at 390 is substituted by Pro"

FT Misc-difference 7 /note= "ATIII.NI35A Ala at 391 is substituted by Asn"

XX WO9958098-A2.

```
Q94D52
ID Q94D52 PRELIMINARY; PRT; 549 AA.
AC Q94D52;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE P0712E02.3 protein (OSUNB0024F06.13 protein).
GN P0712E02.3 OR OSUNB0024F06.13.
OS Oryza sativa (Rice), and
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaeae; Oryza.
OX NCBI_TaxID=4530, 39947;
RN [1]
SEQUENCE FROM N.A.
RP SPECIES=O.sativa; STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
clone:P0712E02.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
SEQUENCE FROM N.A.
RC SPECIES=O.sativa (japonica cultivar-group); STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, BAC
clone:OSUNB0024F06.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF003492; BAB61840.1; -.
DR EMBL; AF003341; BAC01235.1; -.
DR Gramene; Q94D52; -.
SQ SEQUENCE 549 AA; 58838 MW; C32FE2EFD3A8907 CRC64;
Query Match 80.9%; Score 38; DB 10; Length 549;
Best Local Similarity 77.8%; Pred. No. 2e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 STAEGGGGR 9
DB 105 SPADGGGGR 113
| | | | |
| | | | |
RESULT 14
Q8RYV1
ID Q8RYV1 PRELIMINARY; PRT; 814 AA.
AC Q8RYV1;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative receptor protein kinase.
GN P0047F09.12.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
clone:P0047F09.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
CC 1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; AF004145; BAB8275.1; -.
DR Gramene; Q8RYV1; -.
DR InterPro; IPR001480; B_lectin.
DR InterPro; IPR006210; IEGF.
DR InterPro; IPR000719; Prot kinase.
DR InterPro; IPR002290; Ser Thr kinase.
DR InterPro; IPR000858; Slacut Glycop.
DR Pfam; PF01453; Agglutinin; 1.
DR Pfam; PF00069; pkinase; 1.
```

```
DR Pfam; PF00954; S_locus glycop; 1.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00108; B_lectin; 1.
DR SMART; SM00181; EGF; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; Kinase; Receptor; Serine/threonine-protein kinase;
KW Transferase.
SQ SEQUENCE 814 AA; 89409 MW; 4E1B57D381D15628 CRC64;
Query Match 80.9%; Score 38; DB 10; Length 814;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 TAEGGGG 8
DB 607 TAEGGGG 613
| | | | |
| | | | |
RESULT 15
Q9KZNO
ID Q9KZNO PRELIMINARY; PRT; 943 AA.
AC Q9KZNO;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative large secreted protein.
GN SC01402 OR SC1A8A.22C.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=A3(2);
RA Harris D.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96 (1996).
RN [4]
SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinovitch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2).";
RL Nature 417:141-147 (2002).
DR EMBL; AL939108; CAB88898.1; -.
DR InterPro; IPR003594; ATPbind_Atpase.
DR InterPro; IPR003650; HAMP.
DR Pfam; PF00672; HAMP; 1.
DR Pfam; PF02518; HATPase_c; 1.
```

GN ZEP1.
OS Datisca glomerata (Durango root).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Eurosid 1; Cucurbitales; Datisceae; Datisca.
OX NCBI_TaxID=34297;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Root nodules;
RA Okubara P.A., Berry A.M.;
RT "An mRNA (Accession No. AF119050) from Symbiotic Root Nodules of
RT Datisca glomerata with Homology to Zinc Finger Transcription Factor
RT Genes (PGR99-043).";
RL Plant Physiol. 119:1149-1149(1999).
DR EMBL; AF119050; AAD26942.1; -;
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; Zf-C2H2; 2.
DR SMART; SM00355; Znf_C2H2; 2.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 2.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 2.
KW Metal-binding; Zinc; Zinc-finger.
SQ SEQUENCE 247 AA; 26797 MW; 503C90F9425DB00F CRC64;

Query Match 83.0%; Score 39; DB 10; Length 247;
Best Local Similarity 77.8%; Pred. No. 59;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 STAEAGGGG 9
Db 135 TTASGGGG 143
||| |||||

RESULT 6
Q8H7P8 PRELIMINARY; PRT; 758 AA.
AC Q8H7P8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative DEAD/DEAH box RNA helicase protein.
GN Q01384D03.13.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA Wing R.A., Yu Y., Soderlund C., Kim H.-R., Rambo T., Saski C.,
RA Currie J., Collura K.;
RT "Rice Genomic Sequence."
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; ACL11930; AAN62787.1; -;
KW Helicase.
SQ SEQUENCE 758 AA; 83912 MW; D2AE5BECF675B817 CRC64;

Query Match 83.0%; Score 39; DB 10; Length 758;
Best Local Similarity 87.5%; Pred. No. 1.9e-02;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 STAEAGGG 8
Db 103 ATAEGGG 110
||| |||||

RESULT 7
Q8L406 PRELIMINARY; PRT; 85 AA.
AC Q8L406;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE B1103C09.13 protein (P0451D05.21 protein).
GN B1103C09.13 OR P0451D05.21.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, BAC
RT clone:B1103C09.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, PAC
RT clone:P0451D05.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003333; BAB91823.1; -;
DR EMBL; AP003253; BAB92319.1; -;
DR Gramene; Q8L406; -;
SQ SEQUENCE 85 AA; 8532 MW; CA9015F53AD1A284 CRC64;

Query Match 80.9%; Score 38; DB 10; Length 85;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 AEGGGG 9
Db 35 AEGGGG 41
||| |||||

RESULT 8
Q8LN49 PRELIMINARY; PRT; 276 AA.
AC Q8LN49;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN P0031G09.2.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Buell C.R., Yuan Q., Ouyang S., Liu J., Gansberger K., Jones K.M.,
RA Overton II L.L., Teitlin T., Kim M.M., Bera J.J., Jin S.S.,
RA Padrosh D.W., Tallon L.J., Koo H., Zismann V., Hsiao J., Blunt S.,
RA Vanaken S.S., Riedmuller S.B., Utterback T.T., Feldblyum T.V.,
RA Yang Q.Q., Haas B.J., Suh B.B., Peterson J.J., Quackenbush J.,
RA White O., Salzberg S.L., Fraser C.M.;
RT "Oryza sativa chromosome 10 P0031G09 genomic sequence."
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC092211; AAN05004.1; -;
DR Gramene; Q8LN49; -;
KW Hypothetical protein.
SQ SEQUENCE 276 AA; 30133 MW; A1C8C1939F32DF2B CRC64;

Query Match 80.9%; Score 38; DB 10; Length 276;
Best Local Similarity 87.5%; Pred. No. 98;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TAEGGGG 9
Db 104 TASGGGG 111
||| |||||

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 18:03:30 ; Search time 20.3333 Seconds
(without alignments)
114.220 Million cell updates/sec

Title: US-10-014-658-13
Perfect score: 47
Sequence: 1 STRBGGGGR 9

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mic.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_protist.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_virus.*
16: sp_bacterioph.*
17: sp_archaea.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40	85.1	406	15	Q9RDC6 streptomyces
2	40	85.1	549	13	Q9PWF4 Gallus gall
3	39	83.0	137	11	Q8C6J8 mus musculus
4	39	83.0	198	10	Q9FRA0 oryza sativ
5	39	83.0	247	10	Q9XEU0 datisca glo
6	39	83.0	758	10	Q8H7P8 oryza sativ
7	38	80.9	85	10	Q8L406 oryza sativ
8	38	80.9	276	10	Q8L4A9 oryza sativ
9	38	80.9	298	10	Q9FP02 oryza sativ
10	38	80.9	362	10	Q8LNU4 oryza sativ
11	38	80.9	383	10	Q9ARP7 oryza sativ
12	38	80.9	389	5	Q9U0W5 leishmania
13	38	80.9	549	10	Q94D52 oryza sativ
14	38	80.9	814	10	Q8RYV1 oryza sativ
15	38	80.9	943	16	Q9KZNO streptomyces
16	38	80.9	1078	15	Q9A3J0 caulobacter

ALIGNMENTS

RESULT 1

ID	Q9RDC6	PRELIMINARY;	PRT;	406 AA.
AC	Q9RDC6;			
DT	01-MAY-2000 (TREMREL1.13, Created)			
DT	01-MAY-2000 (TREMREL1.13, Last sequence update)			
DT	01-MAR-2003 (TREMREL1.23, Last annotation update)			
DE	Hypothetical protein SCO2497.			
GN	SCO2497 OR SCC82.03C.			
OS	Streptomyces coelicolor.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Streptomycineae; Streptomycetaceae; Streptomycetes.			
OX	NCBI_TaxID=1902;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=A3(2);			
RA	Oliver K., Harris D.;			
RL	Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=A3(2);			
RA	Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;			
RL	Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=A3(2);			
RX	MEDLINE=97000351; PubMed=8843436;			
RA	Redenbach M., Kieser H.M., Denapate D., Eichner A., Cullum J.;			
RA	Kinashi H., Hopwood D.A.;			
RT	"A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome.";			
RL	Mol. Microbiol. 21:77-96(1996).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=A3(2) / M145;			
RX	MEDLINE=21996410; PubMed=12000953;			
RA	Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,			
RA	Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,			
RA	Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,			

Q9RYK2 deinothecium
Q9I3U4 pseudomonas
Q42181 fugu rubrip
Q81EX6 arabidopsis
Q23263 arabidopsis
Q81LP5 oryza sativ
Q81N52 oryza sativ
Q81N73 drosophila
Q9AB13 caulobacter
Q8NDY5 leishmania
Q8MPA9 dictyosteli
Q8H321 oryza sativ
Q8S7Q6 oryza sativ
Q8S7Q6 oryza sativ
Q8S7Q6 oryza sativ
Q9VXK4 drosophila
Q8MS21 dictyosteli
Q8S3Y1 oryza sativ
Q9FPC3 oryza sativ
Q61095 trypanosoma
Q45059 corynebacte
Q923V3 pseudomonas
Q8QK55 ectocarpus
Q947T6 oryza sativ
Q8GR24 oryza sativ
Q32735 agrobacteri
Q9FWE2 oryza sativ
Q9S751 oryza sativ
Q9W182 drosophila
Q9X5B2 neisseria m

```

RT "Temperature acclimation induces light meromyosin isoforms with
RL different primary structures in carp fast skeletal muscle.";
CC Biochem. Biophys. Res. Commun. 208:118-125(1995).
CC
CC -!- FUNCTION: MUSCLE CONTRACTION.
CC
CC -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC-2)
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-1).
CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -!- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEROMYOSIN (LMM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).
CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
CC
CC -!- SIMILARITY: Contains 1 IQ domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; D89992; BAA22069.1; -
CC DR EMBL; D50476; BAA09069.1; -
CC DR EMBL; D43700; BAA07802.1; -
CC DR PIR; I50496; I50496.
CC DR HSP; P13538; 2MYS.
CC DR InterPro; IPR000048; IQ_region.
CC DR InterPro; IPR001609; myosin_head.
CC DR InterPro; IPR004009; Myosin_N.
CC DR InterPro; IPR002928; Myosin_tail.
CC DR Pfam; PF00612; IQ; 1.
CC DR Pfam; PF00063; myosin_head; 1.
CC DR Pfam; PF02736; Myosin_N; 1.
CC DR Pfam; PF01576; Myosin_tail; 1.
CC DR PRINTS; PR00193; MYOSINHEAV.
CC DR ProDom; PD000355; myosin_head; 1.
CC DR SMART; SM00015; IQ; 1.
CC DR SMART; SM00242; MYSC; 1.
CC DR PROSITE; PS00096; IQ; 1.
CC KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
CC ATP-binding; Methylation; Alkylation; Calmodulin-binding;
CC Multigene family.
CC
CC FT DOMAIN 1 781 MYOSIN HEAD-LIKE.
CC FT DOMAIN 782 811 IQ.
CC FT DOMAIN 812 839 HINGE.
CC FT DOMAIN 840 1935 COILED COIL (POTENTIAL).
CC FT NP_BIND 178 185 ATP (POTENTIAL).
CC FT DOMAIN 659 681 ACTIN-BINDING.
CC FT DOMAIN 761 775 ACTIN-BINDING.
CC FT MOD_RES 129 139 METHYLATION (TRP-) (POTENTIAL).
CC FT MOD_RES 696 696 ALKYLATION (SH-1).
CC FT MOD_RES 706 706 ALKYLATION (SH-2).
CC SQ SEQUENCE 1935 AA; 221599 MW; 9A1244B67D63C83B CRC64;
CC
CC Query Match 74.5%; Score 35; DB 1; Length 1935;
CC Best Local Similarity 85.7%; Pred. No. 3.3e+02;
CC Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
CC
CC QY 3 AEGGGGR 9
CC
CC Db 630 AEGGGGK 636
CC
CC RESULT 15
CC MYSS_CHICK STANDARD; PRT; 1938 AA.
CC ID MYSS_CHICK

```

```

AC P13538; O13228;
DT 01-JAN-1990 (Rel. 13, Created)
DT 15-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Myosin heavy chain, skeletal muscle, adult.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=White leghorn; TISSUE=Pectoralis muscle;
RA Chao T.H., Bandman E., Moore L.;
RT "Cloning, nucleotide sequence and characterization of a full-length
RT cDNA encoding the myosin heavy chain from adult chicken pectoralis
RT major muscle.";
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-205.
RC TISSUE=Pectoralis muscle;
RX MEDLINE=92041767; PubMed=1939027;
RA Hayashida M., Maita T., Matsuda G.;
RT "The primary structure of skeletal muscle myosin heavy chain: I.
RT Sequence of the amino-terminal 23 kDa fragment.";
RL J. Biochem. 110:54-59(1991).
RN [3]
RP SEQUENCE OF 206-636.
RC TISSUE=Pectoralis muscle;
RX MEDLINE=92041768; PubMed=1939028;
RA Komine Y., Maita T., Matsuda G.;
RT "The primary structure of skeletal muscle myosin heavy chain: II.
RT Sequence of the 50 kDa fragment of subfragment-1.";
RL J. Biochem. 110:60-67(1991).
RN [4]
RP SEQUENCE OF 637-837.
RC TISSUE=Pectoralis muscle;
RX MEDLINE=92041769; PubMed=1939029;
RA Maita T., Miyaniishi T., Matsuzono K., Tanioka Y., Matsuda G.;
RT "The primary structure of skeletal muscle myosin heavy chain: III.
RT Sequence of the 22 kDa fragment and the alignment of the 23 kDa, 50
RT kDa, and 22 kDa fragments.";
RL J. Biochem. 110:68-74(1991).
RN [5]
RP SEQUENCE OF 838-1938.
RC TISSUE=Pectoralis muscle;
RX MEDLINE=92041770; PubMed=1939030;
RA Maita T., Yajima E., Nagata S., Miyaniishi T., Nakayama S., Matsuda G.;
RT "The primary structure of skeletal muscle myosin heavy chain: IV.
RT Sequence of the rod, and the complete 1,938-residue sequence of the
RT heavy chain.";
RL J. Biochem. 110:75-87(1991).
RN [6]
RP PRELIMINARY SEQUENCE OF 1-808.
RX MEDLINE=97092420; PubMed=3487365;
RA Maita T., Hayashida M., Tanioka Y., Komine Y., Matsuda G.;
RT "The primary structure of the myosin head.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:416-420(1987).
RN [7]
RP SEQUENCE OF 842-1270.
RX MEDLINE=90121764; PubMed=2610940;
RA Watanabe B.;
RT "Complete amino-acid sequence of subfragment-2 in adult chicken
RT skeletal muscle myosin.";
RL Biol. Chem. Hoppe-Seyler 370:1027-1034(1989).
RN [8]
RP SEQUENCE OF 852-1108.
RX MEDLINE=89374803; PubMed=2775482;
RA Watanabe B.;
RT "Amino-acid sequence of the short subfragment-2 in adult chicken
RT skeletal muscle myosin.";
RL Biol. Chem. Hoppe-Seyler 370:549-558(1989).
RN [9]

```

DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Sodium/potassium-transporting ATPase alpha-2 chain precursor
DE (EC 3.6.3.9) (Sodium pump 2) (Na+/K+ ATPase 2).
GN ATP1A2 OR KIAA0778.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=90008924; PubMed=2477373;
RA Shull M.M., Pugh D.G., Lingrel J.B.;
RT "Characterization of the human Na,K-ATPase alpha 2 gene and
RT identification of intragenic restriction fragment length
RT polymorphisms.";
RL J. Biol. Chem. 264:17532-17543 (1989).
RN [2]
RN SEQUENCE OF 211-249 FROM N.A.
RP TISSUE=Leukocyte;
RX MEDLINE=87231946; PubMed=3035563;
RA Shull M.M., Lingrel J.B.;
RT "Multiple genes encode the human Na,K-ATPase catalytic subunit.";
RT Proc. Natl. Acad. Sci. U.S.A. 84:4039-4043 (1987).
RN [3]
RN SEQUENCE OF 251-442 FROM N.A.
RP TISSUE=Brain, and Placenta;
RX MEDLINE=87247232; PubMed=3036582;
RA Sverdlov E.D., Monastyrskaya G.S., Broude N.E., Ushkaryov Y.A.,
RA Allikmets R.L., Melkov A.M., Smirnov Y.V., Malyshev I.V.,
RA Dulobova I.S., Petrukhin K.E., Grishin A.V., Kijackin N.I.,
RA Kostina M.B., Sverdlov V.E., Modyanov N.N., Ovchinnikov Y.A.;
RT "The family of human Na,K-ATPase genes. No less than five genes
RT and/or pseudogenes related to the alpha-subunit.";
RT FEBS Lett. 217:275-278 (1987).
RN [4]
RN SEQUENCE OF 1-4 FROM N.A.
RP MEDLINE=89153603; PubMed=2537767;
RA Sverdlov E.D., Bessarab D.A., Malyshev I.V., Petrukhin K.E.,
RA Smirnov Y.V., Ushkaryov Y.A., Monastyrskaya G.S., Broude N.E.,
RA Modyanov N.N.;
RT "Family of human Na,K-ATPase genes. Structure of the putative
RT regulatory region of the alpha+-gene.";
RT FEBS Lett. 244:481-483 (1989).
RN [5]
RN SEQUENCE FROM N.A.
RP TISSUE=Brain;
RX MEDLINE=99087487; PubMed=9872452;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XI.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 5:277-286 (1998).
CC -I- FUNCTION: This is the catalytic component of the active enzyme,
CC which catalyzes the hydrolysis of ATP coupled with the exchange of
CC sodium and potassium ions across the plasma membrane. This action
CC creates the electrochemical gradient of sodium and potassium.
CC Providing the energy for active transport of various nutrients.
CC -I- CATALYTIC ACTIVITY: ATP + H₂O + Na⁺ (in) + K⁺ (out) = ADP +
CC phosphate + Na⁺ (out) + K⁺ (in).
CC -I- SUBUNIT: COMPOSED OF THREE SUBUNITS: ALPHA (CATALYTIC), BETA
CC AND GAMMA.
CC -I- SUBCELLULAR LOCATION: Integral membrane protein.
CC -I- SIMILARITY: Belongs to the cation transport ATPases family (P-type
CC ATPases). Subfamily IIC.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

or send an email to license@isb-sib.ch.
CC -----
CC EMBL; J05096; AAA51797.1; -
DR EMBL; M16795; AAA51799.1; -
DR EMBL; M27578; AAA35575.1; -
DR EMBL; M27571; AAA35575.1; JOINED.
DR EMBL; M27576; AAA35575.1; JOINED.
DR EMBL; Y07494; CAA68793.1; ALT_SEQ.
DR EMBL; AB018321; BAA34498.2; -
DR PIR; A34474; A34474.
DR HSSP; P04191; IEUL.
DR GENE; HGNC:800; ATP1A2.
DR MIM; 182340; -
DR InterPro; IPR001757; ATPase_E1-E2.
DR InterPro; IPR006069; Cation_ATPase.
DR InterPro; IPR006088; Cation_ATPase_C.
DR InterPro; IPR004014; Cation_ATPase_N.
DR InterPro; IPR005834; Hydrolyase.
DR InterPro; IPR005775; Na/K_ATPase_alpha.
DR Pfam; PF00689; Cation_ATPase_N; 1.
DR Pfam; PF00122; E1-E2_ATPase; 1.
DR Pfam; PF00702; Hydrolyase; 1.
DR PRINTS; PR00119; CATIONPASE.
DR PRINTS; PR00121; NAKATPASE.
DR TIGRFAMs; TIGR01106; ATPase-IIC_X-K; 1.
DR TIGRFAMs; TIGR01494; ATPase_P-type; 5.
DR PROSITE; PS00154; ATPase_E1-E2; 1.
KW Hydrolyase; Sodium/potassium transport; Transmembrane; Phosphorylation;
KW Magnesium; Metal-binding; ATP-binding; Multigene family.
FT PROPEP 1 5 BY SIMILARITY.
FT CHAIN 6 1020 SODIUM/POTASSIUM-TRANSPORTING ATPASE
FT DOMAIN 6 85 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 86 106 POTENTIAL.
FT DOMAIN 107 129 LUMENAL (POTENTIAL).
FT TRANSMEM 130 150 POTENTIAL.
FT DOMAIN 151 286 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 287 306 POTENTIAL.
FT DOMAIN 307 318 LUMENAL (POTENTIAL).
FT TRANSMEM 319 336 POTENTIAL.
FT DOMAIN 337 769 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 770 789 POTENTIAL.
FT DOMAIN 790 799 LUMENAL (POTENTIAL).
FT TRANSMEM 800 820 POTENTIAL.
FT DOMAIN 821 840 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 841 863 POTENTIAL.
FT DOMAIN 864 915 LUMENAL (POTENTIAL).
FT TRANSMEM 916 935 POTENTIAL.
FT DOMAIN 936 948 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 949 967 POTENTIAL.
FT DOMAIN 968 982 LUMENAL (POTENTIAL).
FT TRANSMEM 983 1003 POTENTIAL.
FT DOMAIN 1004 1020 CYTOPLASMIC (POTENTIAL).
FT MOD_RES 374 374 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 940 940 PHOSPHORYLATION (BY PRA) (BY SIMILARITY).
FT BINDING 80 82 BINDING OF PHOSPHOINOSITIDE-3 KINASE (BY
FT SIMILARITY).
FT METAL 714 714 MAGNESIUM (BY SIMILARITY).
FT METAL 718 718 MAGNESIUM (BY SIMILARITY).
SQ SEQUENCE 1020 AA; 112265 MW; AFDDBEA94FFB4FC3 CRC64;
Query Match 74.5%; Score 35; DB 1; Length 1020;
Best Local Similarity 66.7%; Pred. No. 1.7e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Qy 1 STAEKGGGR 9
Db 14 TTAENGCGK 22
RESULT 13
A1A2_RAT

```
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93185339; PubMed=1363401;
RA Logan C., Hanks M.C., Noble-Topham S., Nallainathan D.,
RA Provart N.J., Joyner A.L.;
RT "Cloning and sequence comparison of the mouse, human, and chicken
engrafted genes reveal potential functional domains and regulatory
regions."
RT Dev. Genet. 13:345-358(1992).
RL [2]
RP REVISION TO 229.
RA Logan C., Hanks M.C., Noble-Topham S., Nallainathan D., Provart N.J.,
RA Joyner A.L.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 230-333 FROM N.A.
RX MEDLINE=89233109; PubMed=2565873;
RA Poole S.J., Law M.L., Kao F.T., Lau Y.F.;
RT "Isolation and chromosomal localization of the human En-2 gene."
RL Genomics 4:225-231(1989).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE ENGRAILED HOMEOBOX FAMILY.
CC -!- SIMILARITY: Contains 1 homeobox domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
DR EMBL; L12701; AAAS3504.2; -
DR EMBL; L12700; AAAS3504.2; JOINED.
DR EMBL; J03056; AAP8670.1; -
DR F01; E48423; E48423.
DR HSP; P02836; 3HDD.
DR TRANSFAC; T02019; -
DR Genew; HGNC:3343; EN2.
DR MIM; 131310; -
DR GO; GO:000725; P:development; TAS.
DR InterPro; IPR000747; Engrailed.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR000047; HTH_lambdarepressor.
DR Pfam; PF00046; homeobox; 1.
DR PRINTS; PR00026; ENGRAILED.
DR PRINTS; PR00024; HOMEOBOX.
DR PRINTS; PR00031; HTHREPRESSOR.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEOBOX_1; 1.
DR PROSITE; PS50071; HOMEOBOX_2; 1.
DR PROSITE; PS00033; ENGRAILED; 1.
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT DNA_BIND 244 303 HOMEOBOX.
SQ SEQUENCE 333 AA; 34210 MW; ACF5399E383D6257 CRC64;

Query Match 74.5%; Score 35; DB 1; Length 333;
Best Local Similarity 87.5%; Pred. No. 57;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 STAEGGG 8
Db 107 SGAEGGG 114

RESULT 9
VL2 HPV63
ID _VL2_HPV63 STANDARD; PRT; 504 AA.

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93185339; PubMed=1363401;
RA Logan C., Hanks M.C., Noble-Topham S., Nallainathan D.,
RA Provart N.J., Joyner A.L.;
RT "Cloning and sequence comparison of the mouse, human, and chicken
engrafted genes reveal potential functional domains and regulatory
regions."
RT Dev. Genet. 13:345-358(1992).
RL [2]
RP REVISION TO 229.
RA Logan C., Hanks M.C., Noble-Topham S., Nallainathan D., Provart N.J.,
RA Joyner A.L.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 230-333 FROM N.A.
RX MEDLINE=89233109; PubMed=2565873;
RA Poole S.J., Law M.L., Kao F.T., Lau Y.F.;
RT "Isolation and chromosomal localization of the human En-2 gene."
RL Genomics 4:225-231(1989).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE ENGRAILED HOMEOBOX FAMILY.
CC -!- SIMILARITY: Contains 1 homeobox domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
DR EMBL; L12701; AAAS3504.2; -
DR EMBL; L12700; AAAS3504.2; JOINED.
DR EMBL; J03056; AAP8670.1; -
DR F01; E48423; E48423.
DR HSP; P02836; 3HDD.
DR TRANSFAC; T02019; -
DR Genew; HGNC:3343; EN2.
DR MIM; 131310; -
DR GO; GO:000725; P:development; TAS.
DR InterPro; IPR000747; Engrailed.
DR InterPro; IPR001356; Homeobox.
DR InterPro; IPR000047; HTH_lambdarepressor.
DR Pfam; PF00046; homeobox; 1.
DR PRINTS; PR00026; ENGRAILED.
DR PRINTS; PR00024; HOMEOBOX.
DR PRINTS; PR00031; HTHREPRESSOR.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEOBOX_1; 1.
DR PROSITE; PS50071; HOMEOBOX_2; 1.
DR PROSITE; PS00033; ENGRAILED; 1.
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT DNA_BIND 244 303 HOMEOBOX.
SQ SEQUENCE 333 AA; 34210 MW; ACF5399E383D6257 CRC64;

Query Match 74.5%; Score 35; DB 1; Length 333;
Best Local Similarity 87.5%; Pred. No. 57;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 STAEGGG 8
Db 107 SGAEGGG 114

RESULT 9
VL2 HPV63
ID _VL2_HPV63 STANDARD; PRT; 504 AA.
```

```
AC Q07863;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Minor capsid protein L2.
GN Human papillomavirus type 63.
OS Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=28311;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93276568; PubMed=8389082;
RA Egawa K., Delius H., Matsukura T., Kawashima M., de Villiers E.M.;
RT "Two novel types of human papillomavirus, HPV 63 and HPV 65:
comparisons of their clinical and histological features and DNA
sequences to other HPV types."
RT Virology 194:789-799(1993).
RL [2]
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
DR EMBL; X70828; CAAS0169.1; -
DR InterPro; IPR000784; Late_L2.
DR Pfam; PF00513; late_protein_L2; 1.
KW Coat protein; Late protein.
SQ SEQUENCE 504 AA; 54826 MW; F74758B3C84CDB6B CRC64;

Query Match 74.5%; Score 35; DB 1; Length 504;
Best Local Similarity 75.0%; Pred. No. 86;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 TAEGGGG 9
Db 58 TGKGGGG 65

RESULT 10
KF3C RAT
ID _KF3C_RAT STANDARD; PRT; 796 AA.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98155212; PubMed=94871132;
RA Muresan V., Abramson T., Lyass A., Winter D., Porro E., Hong F.,
RA Chamberlin N.L., Schnapp B.J.;
RT "KIF3C and KIF3A form a novel neuronal heteromeric kinesin that
associates with membrane vesicles."
RL Mol. Biol. Cell 9:637-652(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99023472; PubMed=9808286;
RA Faure K., Gruber D., Bulinski J.C.;
RT "Identification of kinesin-like molecules in myogenic cells."
RL Eur. J. Cell Biol. 77:27-34(1998).
CC -!- FUNCTION: MICROTUBULE-BASED ANTEROGRADE TRANSLATOR FOR
MEMBRANOUS ORGANELLES.
CC -!- SUBUNIT: HETERODIMER OF KIF3A AND KIF3C.
CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. KINESIN
```


CC EMBL; AB012523; AAM43240.1; --
DR HAMAP; MF_00333; --; 1.
DR InterPro; IPR001260; Coprogen oxidase.
DR Pfam; PF01218; Coprogen oxidase; 1.
DR PRINTS; PR00073; COPROGENOXIDASE.
DR PROSITE; PS01021; COPROGEN OXIDASE; 1.
KW Porphyrin biosynthesis; Oxidoreductase; Iron; Complete proteome.
SQ SEQUENCE 299 AA; 34068 MW; 3CAEFDCC50FA1B3 CRC64;

Query Match 80.9%; Score 38; DB 1; Length 299;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 AEGGGGR 9
DB 39 AEGGGGR 45

RESULT 4
CO3_LAMJA STANDARD; PRT; 1673 AA.
ID CO3_LAMJA STANDARD; PRT; 1673 AA.
AC Q00685; Rel. 33, Created
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Complement C3 precursor [Contains: C3a anaphylatoxin] (Fragment).
GN C3.
OS Lampetra japonica (Japanese lamprey) (Entosphenus japonicus).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC Petromyzontiformes; Petromyzontidae; Lethenteron.
OX NCBI_TaxID=94989;
RN [1]
RS SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=92251197; PubMed=1578150;
RA Nonaka M., Takahashi M.;
RT "Complete complementary DNA sequence of the third component of
RT complement of lamprey. Implication for the evolution of thioester
RT containing proteins.";
RL J. Immunol. 148:3290-3295(1992).
CC -1- FUNCTION: C3 PLAYS A CENTRAL ROLE IN THE ACTIVATION OF THE
CC COMPLEMENT SYSTEM. AFTER ACTIVATION (C3B), IT CAN BIND COVALENTLY,
CC VIA ITS REACTIVE THIOLESTER, TO CELL SURFACE CARBOHYDRATES OR
CC IMMUNE AGGREGATES. CYCLOSTOMATES C3 APPEARS TO REPRESENT THE
CC COMMON ANCESTOR OF MAMMALIAN C3 AND C4, SHOWING SIMILARITIES TO
CC BOTH PROTEINS.
CC -1- SIMILARITY: TO C4, C5 AND ALPHA-2-MACROGLOBULIN.
CC -1- SIMILARITY: Contains 1 anaphylatoxin-like domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D10087; BAA00983.1; --
DR PIR; I50806; I50806.
DR HSP; P01024; 1C3D.
DR InterPro; IPR002890; A2M N.
DR InterPro; IPR000020; Anaphylatoxin.
DR InterPro; IPR001599; MacroglobulinA2.
DR InterPro; IPR001134; Netrin_C.
DR Pfam; PF00207; A2M; 1.
DR Pfam; PF01835; A2M_N; 1.
DR Pfam; PF01821; ANATO; 1.
DR Pfam; PF01759; NTR; 1.
DR ProDom; PD003264; Anaphylatoxin; 1.
DR SMART; SMO0104; ANATO; 1.
DR SMART; SMO0643; C345C; 1.
DR PROSITE; PS00477; ALPHA_2_MACROGLOBULIN; 1.

DR PROSITE; PS01177; ANAPHYLATOXIN 1; 1.
DR PROSITE; PS01178; ANAPHYLATOXIN 2; 1.
KW Complement pathway; Plasma; Inflammatory response; Glycoprotein;
KW Signal; Thioester bond.
FT NON_TER 1 1
FT SIGNAL <1 13 POTENTIAL.
FT CHAIN 14 1673 COMPLEMENT C3.
FT CHAIN 14 653 BETA CHAIN (BY SIMILARITY).
FT CHAIN 657 1375 ALPHA CHAIN (BY SIMILARITY).
FT CHAIN 1379 1673 GAMMA CHAIN (BY SIMILARITY).
FT PEPTIDE 657 732 C3A ANAPHYLATOXIN-LIKE.
FT DOMAIN 678 713 C3A ANAPHYLATOXIN (BY SIMILARITY).
FT DISULFID 678 705 BY SIMILARITY.
FT DISULFID 679 712 BY SIMILARITY.
FT DISULFID 692 713 BY SIMILARITY.
FT CROSSLINK 986 989 Iso-glutamyl cysteine thioester (Cys-Gln)
FT (By similarity).
SQ SEQUENCE 1673 AA; 187767 MW; 4B5FC756B123A506 CRC64;

Query Match 78.7%; Score 37; DB 1; Length 1673;
Best Local Similarity 87.5%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TAEGGGGR 9
DB 614 TAEGGGGR 621

RESULT 5
PHT4_PSEPU STANDARD; PRT; 410 AA.
ID PHT4_PSEPU STANDARD; PRT; 410 AA.
AC Q05184;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Putative 4,5-dihydroxyphthalate dehydrogenase (EC 1.1.1.1)
DE (DHP dehydrogenase).
GN PHT4.
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=303;
RN [1]
RS SEQUENCE FROM N.A.
RC STRAIN=NMH102-2;
RA Nomura Y., Nakagawa M., Ogawa N., Harashima S., Oshima Y.;
RT "Genes in PHT plasmid encoding the initial degradation pathway of
RT phthalate in Pseudomonas putida.";
RL J. Ferment. Bioeng. 74:333-344(1992).
CC -1- FUNCTION: TRANSFORMS 4,5-DIHYDRO-4,5-DIHYDROXYPHTHALATE TO
CC 4,5-DIHYDROXYPHTHALATE.
CC -1- PATHWAY: Phthalate degradation; second step.
CC -1- INDUCTION: INDUCED BY PHTHALATE AND REPRESSED BY GLUCOSE.
CC -1- SIMILARITY: TO ALCALIGENES SP. 1-CARBOXY-3-CHLORO-3,4-
CC DIHYDROXYCYCLOHEXA-1,5-DIENE DEHYDROGENASE.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D13229; BAA03512.1; --
DR InterPro; IPR000693; GFO_IDH_MoCA.
DR InterPro; IPR004104; GFO_IDH_MoCA_C.
DR Pfam; PF01408; GFO_IDH_MoCA; 1.
DR Pfam; PF02894; GFO_IDH_MoCA_C; 1.
KW Plasmid; Oxidoreductase; Aromatic hydrocarbons catabolism.
SQ SEQUENCE 410 AA; 45644 MW; 0D1A95B019DCACIA CRC64;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 16:54:27 ; Search time 4.66667 Seconds
(without alignments)
90.694 Million cell updates/sec

Title: US-10-014-658-13

Perfect score: 47

Sequence: 1 STABEGGGR 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	87.2	255	1 MTN MYCTU	Q10889 mycobacteri
2	40	85.1	249	1 RELB CHICK	P51509 gallus gall
3	38	80.9	299	1 HEM6_XANCP	Q06300 xanthomonas
4	37	78.7	1673	1 CO3 LAMJA	Q00685 lampetra ja
5	26	76.6	410	1 PH74_PSEPU	Q05184 pseudomonas
6	26	76.6	450	1 SRP_DROME	Q24535 drosophila
7	36	76.6	1441	1 CPSA_MOUSE	Q98944 mus musculu
8	35	74.5	333	1 HME2_HUMAN	P19622 homo sapien
9	35	74.5	504	1 VL2_HPV63	Q07863 human papil
10	35	74.5	796	1 KF3C_RAT	O55165 rattus norv
11	35	74.5	1017	1 ALA2_CHICK	P24797 gallus gall
12	35	74.5	1020	1 ALA2_HUMAN	P05993 homo sapien
13	35	74.5	1020	1 ALA2_RAT	P06686 rattus norv
14	35	74.5	1935	1 MYSS_CIPCA	Q90339 cyprinus ca
15	35	74.5	1938	1 MYSS_CHICK	P13538 gallus gall
16	35	74.5	1939	1 MYH4_HUMAN	Q9Y623 homo sapien
17	34	72.3	299	1 HEM6_XANAC	Q8pf76 xanthomonas
18	34	72.3	305	1 HEM6_XYLFA	Q9phc7 xyella fas
19	34	72.3	309	1 HEM6_VERPE	Q8zcf9 versinia pe
20	34	72.3	324	1 HME2_MOUSE	P09066 mus musculu
21	34	72.3	340	1 HEME_STNY3	P72848 synechocyst
22	34	72.3	456	1 FXD1_MOUSE	Q61343 mus musculu
23	34	72.3	507	1 UL17_EBV	P03222 Epstein-Bar
24	34	72.3	533	1 EGRI_MOUSE	P08046 mus musculu
25	34	72.3	542	1 YQ16_CABEL	Q09279 caenorhabdi
26	34	72.3	578	1 CES6_HUMAN	Q9bxq6 homo sapien
27	34	72.3	585	1 ESR1_ORENI	Q9wh33 orochromis
28	34	72.3	620	1 ESR1_ORILA	P50241 oryzias lat
29	34	72.3	707	1 RHO_STRLI	P52157 streptomyce
30	34	72.3	741	1 FIB4_CHICK	P14448 gallus gall
31	34	72.3	3149	1 PFDB_EBV	P03186 Epstein-Bar
32	33	70.2	120	1 PFDB_METKA	Q8tyc7 methanopyru
33	33	70.2	141	1 DUT_CHVP1	Q41033 paramecium

RESULT 1

ID	MTN MYCTU	STANDARD	PRT	255 AA
AC	Q10889			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	MTA/SAH nucleosidase [Includes: 5'-methylthioadenosine nucleosidase (EC 3.2.2.9)]			
DE	(EC 3.2.2.16); S-adenosylhomocysteine nucleosidase (EC 3.2.2.9)]			
GN	MTN OR PFS OR RV0091 OR MT0100 OR MIC251.10.			
OS	Mycobacterium tuberculosis.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Corynebacterineae; Mycobacteriaceae; Mycobacterium.			
OX	NCBI_TaxID=1773;			
FN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=H37RV;			
RC	MEDLINE=98295987; PubMed=9634230;			
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,			
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,			
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,			
RA	Davies R., Delvin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,			
RA	Hornsby T., Jagals K., Krogh A., McLean J., Moule S., Murphy L.,			
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,			
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,			
RA	Sulston J.E., Taylor K., Whitehead S., Barrell B.G.,			
RT	"Deciphering the biology of Mycobacterium tuberculosis from the			
RT	complete genome sequence."			
RL	Nature 393:537-544 (1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CDC 1551 / Oshkosh;			
RC	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,			
RA	Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,			
RA	Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,			
RA	Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,			
RA	Bisshai W.,			
RT	"Whole genome comparison of Mycobacterium tuberculosis clinical and			
RT	laboratory strains."			
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.			
CC	-1- FUNCTION: RESPONSIBLE FOR CLEAVAGE OF THE GLYCOSIDIC BOND IN BOTH			
CC	5'-METHYLTHIOADENOSINE (MTA) AND S-ADENOSYLHOMOCYSTEINE (SAH) (BY			
CC	SIMILARITY).			
CC	-1- CATALYTIC ACTIVITY: S-adenosyl-L-homocysteine + H(2)O = adenine +			
CC	S-D-ribosyl-L-homocysteine.			
CC	-1- CATALYTIC ACTIVITY: Methylthioadenosine + H(2)O = adenine + 5-			
CC	methylthio-D-ribose.			
CC	-1- SIMILARITY: BELONGS TO THE MTN FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			

ALIGNMENTS

ID	MTN MYCTU	STANDARD	PRT	255 AA
AC	Q10889			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	MTA/SAH nucleosidase [Includes: 5'-methylthioadenosine nucleosidase (EC 3.2.2.9)]			
DE	(EC 3.2.2.16); S-adenosylhomocysteine nucleosidase (EC 3.2.2.9)]			
GN	MTN OR PFS OR RV0091 OR MT0100 OR MIC251.10.			
OS	Mycobacterium tuberculosis.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Corynebacterineae; Mycobacteriaceae; Mycobacterium.			
OX	NCBI_TaxID=1773;			
FN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=H37RV;			
RC	MEDLINE=98295987; PubMed=9634230;			
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,			
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,			
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,			
RA	Davies R., Delvin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,			
RA	Hornsby T., Jagals K., Krogh A., McLean J., Moule S., Murphy L.,			
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,			
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,			
RA	Sulston J.E., Taylor K., Whitehead S., Barrell B.G.,			
RT	"Deciphering the biology of Mycobacterium tuberculosis from the			
RT	complete genome sequence."			
RL	Nature 393:537-544 (1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CDC 1551 / Oshkosh;			
RC	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,			
RA	Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,			
RA	Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,			
RA	Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,			
RA	Bisshai W.,			
RT	"Whole genome comparison of Mycobacterium tuberculosis clinical and			
RT	laboratory strains."			
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.			
CC	-1- FUNCTION: RESPONSIBLE FOR CLEAVAGE OF THE GLYCOSIDIC BOND IN BOTH			
CC	5'-METHYLTHIOADENOSINE (MTA) AND S-ADENOSYLHOMOCYSTEINE (SAH) (BY			
CC	SIMILARITY).			
CC	-1- CATALYTIC ACTIVITY: S-adenosyl-L-homocysteine + H(2)O = adenine +			
CC	S-D-ribosyl-L-homocysteine.			
CC	-1- CATALYTIC ACTIVITY: Methylthioadenosine + H(2)O = adenine + 5-			
CC	methylthio-D-ribose.			
CC	-1- SIMILARITY: BELONGS TO THE MTN FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			

A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-687 <STO>
A:Cross-references: GB:AE005673; NID:g13421339; PIDN:AAK22201.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC0214

Query Match 78.7%; Score 37; DB 2; Length 687;
Best Local Similarity 87.5%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TAEGGGGR 9
DB 240 TAEGGAGR 247
|||||

RESULT 8
150806
complement component C3 - Japanese lamprey (fragment)
C:Species: Lampetra japonica (Japanese lamprey)
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 16-Jul-1999
C:Accession: I50806
R:Nonaka, M.; Takahashi, M.
J. Immunol. 148, 3290-3295, 1992
A:Title: Complete complementary DNA sequence of the third component of complement of lamprey
A:Reference number: I50806; MUID:92251197; PMID:1578150
A:Accession: I50806
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1673 <NON>
A:Cross-references: GB:D10087; NID:g222887; PIDN:BA00983.1; PID:g222888
C:Superfamily: alpha-2-macroglobulin

Query Match 78.7%; Score 37; DB 2; Length 1673;
Best Local Similarity 87.5%; Pred. No. 3.1e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TAEGGGGR 9
DB 614 TAEGGAGR 621
|||||

RESULT 9
S60674
hypothetical protein B - Corynebacterium glutamicum plasmid pGAI
C:Species: Corynebacterium glutamicum
C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 22-Oct-1999
C:Accession: S60674
R:Nesvera, J.; Patek, M.; Hochmannova, J.; Abzhomova, Z.
submitted to the EMBL Data Library, August 1995
A:Description: Complete nucleotide sequence of the cryptic plasmid pGAI from Corynebacterium glutamicum
A:Reference number: S60673
A:Accession: S60674
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-266 <NES>
A:Cross-references: EMBL:X90817; NID:g951006; PIDN:CAA62329.1; PID:g951008
C:Genetics:
A:Genome: plasmid pGAI

Query Match 76.6%; Score 36; DB 2; Length 266;
Best Local Similarity 75.0%; Pred. No. 80;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 TAEGGGGR 9
DB 50 TAKGGGK 57
|||||

RESULT 10
AH3175
ABC transporter, membrane spanning protein attF [imported] - Agrobacterium tumefaciens
C:Species: Agrobacterium tumefaciens

A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-687 <STO>
A:Cross-references: GB:AE005673; NID:g13421339; PIDN:AAK22201.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC0214

Query Match 78.7%; Score 37; DB 2; Length 687;
Best Local Similarity 87.5%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TAEGGGGR 9
DB 240 TAEGGAGR 247
|||||

RESULT 8
150806
complement component C3 - Japanese lamprey (fragment)
C:Species: Lampetra japonica (Japanese lamprey)
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 16-Jul-1999
C:Accession: I50806
R:Nonaka, M.; Takahashi, M.
J. Immunol. 148, 3290-3295, 1992
A:Title: Complete complementary DNA sequence of the third component of complement of lamprey
A:Reference number: I50806; MUID:92251197; PMID:1578150
A:Accession: I50806
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1673 <NON>
A:Cross-references: GB:D10087; NID:g222887; PIDN:BA00983.1; PID:g222888
C:Superfamily: alpha-2-macroglobulin

Query Match 78.7%; Score 37; DB 2; Length 1673;
Best Local Similarity 87.5%; Pred. No. 3.1e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TAEGGGGR 9
DB 614 TAEGGAGR 621
|||||

RESULT 9
S60674
hypothetical protein B - Corynebacterium glutamicum plasmid pGAI
C:Species: Corynebacterium glutamicum
C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 22-Oct-1999
C:Accession: S60674
R:Nesvera, J.; Patek, M.; Hochmannova, J.; Abzhomova, Z.
submitted to the EMBL Data Library, August 1995
A:Description: Complete nucleotide sequence of the cryptic plasmid pGAI from Corynebacterium glutamicum
A:Reference number: S60673
A:Accession: S60674
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-266 <NES>
A:Cross-references: EMBL:X90817; NID:g951006; PIDN:CAA62329.1; PID:g951008
C:Genetics:
A:Genome: plasmid pGAI

Query Match 76.6%; Score 36; DB 2; Length 266;
Best Local Similarity 75.0%; Pred. No. 80;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 TAEGGGGR 9
DB 50 TAKGGGK 57
|||||

RESULT 10
AH3175
ABC transporter, membrane spanning protein attF [imported] - Agrobacterium tumefaciens
C:Species: Agrobacterium tumefaciens

C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AH3175
R:Wood, D.W.; Seubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.; Erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClal.; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AH3175
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-420 <KUR>
A:Cross-references: GB:AE008687; PIDN:AAL45822.1; PID:g17743561; GSPDB:GN00188
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: attF
A:Genome: plasmid

Query Match 76.6%; Score 36; DB 2; Length 420;
Best Local Similarity 77.8%; Pred. No. 1.2e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 STAEGGGGR 9
DB 356 STAGGGGR 364
|||||

RESULT 11
S42825
serum response factor homolog - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 24-Sep-1998
C:Accession: S42825
R:Grope, J.C.
submitted to the EMBL Data Library, February 1994
A:Reference number: S42825
A:Accession: S42825
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-450 <GRO>
A:Cross-references: EMBL:X77532; NID:g453586; PID:g453587
C:Genetics:
A:Gene: FlyBase:Serf.
A:Cross-references: FlyBase:Fgn0010354
C:Superfamily: serum response factor DNA-binding domain homology
F:166-221/Domain: serum response factor DNA-binding domain homology <SRF>

Query Match 76.6%; Score 36; DB 2; Length 450;
Best Local Similarity 87.5%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 STAEGGGGR 8
DB 353 STAGGGGR 360
|||||

RESULT 12
I54059
Na+/K+-exchanging ATPase (EC 3.6.3.9) alpha-2 chain - rat (fragment)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change 19-Apr-2002
C:Accession: I54059
R:Kawakami, K.; Yagawa, Y.; Nagano, K.
Gene 91, 267-270, 1990
A:Title: Regulation of Na + , K + -ATPases. I. Cloning and analysis of the 5'-flanking region
A:Reference number: I54059; MUID:91007285; PMID:2170235
A:Accession: I54059
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-39 <RES>
A:Cross-references: GB:D90049; NID:g220662; PIDN:BAA14102.1; PID:g220663

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 18:19:54 ; Search time 8.33333 Seconds
(without alignments)
103.862 Million cell updates/sec

Title: US-10-014-658-13

Perfect score: 47

Sequence: 1 STABGGGGR 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168692 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

PIR 76: *
1: Pirl: *
2: Pirl2: *
3: Pirl3: *
4: Pirl4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	87.2	255	2 C70750	hypothetical prote
2	38	80.9	1078	2 D87647	hypothetical prote
3	38	80.9	1091	2 C75585	carboxylase - Dein
4	38	80.9	1095	2 B83471	probable pyruvate
5	38	80.9	1172	2 T49330	cytokinesis inhibi
6	37	78.7	111	2 C71401	hypothetical prote
7	37	78.7	687	2 E87275	TonB-dependent rec
8	37	78.7	1673	2 I50806	complement compone
9	36	76.6	266	2 S60674	hypothetical prote
10	36	76.6	420	2 AH3175	ABC transporter, m
11	36	76.6	450	2 S42825	serum response fac
12	35	74.5	39	2 I54059	Na+/K+-exchanging
13	35	74.5	183	2 E72459	hypothetical prote
14	35	74.5	333	2 E48423	homeotic protein e
15	35	74.5	400	2 T29121	hypothetical prote
16	35	74.5	540	2 E87370	leucine aminopepti
17	35	74.5	596	2 F87297	feruloyl-CoA synth
18	35	74.5	607	2 AG3323	penicillin-binding
19	35	74.5	716	2 S28185	phenylalanine ammo
20	35	74.5	810	2 T48835	lethal(2)denticlel
21	35	74.5	1017	2 A37227	Na+/K+-exchanging
22	35	74.5	1020	2 B24639	Na+/K+-exchanging
23	35	74.5	1020	2 A34474	Na+/K+-exchanging
24	35	74.5	1938	1 JX0178	myosin heavy chain
25	34	72.3	106	2 A34224	hypothetical prote
26	34	72.3	173	2 AG2762	single-strand DNA
27	34	72.3	194	2 F97543	single-strand bind
28	34	72.3	282	2 T13526	hypothetical prote
29	34	72.3	304	2 AC3064	3-hydroxyisobutyr

30 34 72.3 304 2 E98222 probable 3-hydroxy
31 34 72.3 309 2 AG0368 coproporphyrinogen
32 34 72.3 317 2 A11887 coproporphyrinogen
33 34 72.3 319 2 H82858 coproporphyrinogen
34 34 72.3 324 2 D48423 homeotic protein e
35 34 72.3 333 2 S76405 hypothetical prote
36 34 72.3 340 2 S74712 coproporphyrinogen
37 34 72.3 340 2 T20807 hypothetical prote
38 34 72.3 382 2 D72095 conserved hypothet
39 34 72.3 382 2 H85526 CT195 hypothetical
40 34 72.3 461 2 T49693 hypothetical prote
41 34 72.3 488 2 B95332 hypothetical prote
42 34 72.3 507 1 QQBE41 BGLF1 protein - hu
43 34 72.3 533 2 JS0304 developmental cont
44 34 72.3 558 2 G87414 hypothetical prote
45 34 72.3 607 2 T01904 hypothetical prote

ALIGNMENTS

RESULT 1

C70750
hypothetical protein Rv0091 - Mycobacterium tuberculosis (strain H37Rv)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: C70750
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: C70750
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-255 <COL>
A;Cross-references: GB:Z74410; GB:AL123456; NID:G3261600; PIDN:CAA98927.1; PID:G1405762
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv0091
C;Superfamily: Escherichia coli pfs protein

Query Match 87.2%; Score 41; DB 2; Length 255;
Best Local Similarity 88.9%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Cy 1 STABGGGGR 9
Db 152 STABGGGGR 160

RESULT 2

D87647
hypothetical protein CC3214 [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C;Accession: D87647
R;Niernan, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Koloi
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: D87647
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1078 <STO>
A;Cross-references: GB:AE005673; NID:G13424894; PIDN:AAK25176.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC3214

ORGANISM: Streptomyces avermitilis
US-10-156-761-12539
Query Match 76.6%; Score 36; DB 15; Length 726;
Best Local Similarity 87.5%; Pred. No. 8.8e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 STAEGGGG 9
|||:|||||
Db 420 TAVGGGG 427

RESULT 12
US-10-252-408-17
; Sequence 17, Application US/10252408
; Publication No. US20030082736A1
; GENERAL INFORMATION:
; APPLICANT: SMITH, Craig A.
; TITLE OF INVENTION: TUMOR NECROSIS FACTOR-ALPHA AND BETA-RECEPTORS
; FILE REFERENCE: A-71592
; CURRENT APPLICATION NUMBER: US/10/252,408
; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: US/08/406,824
; PRIOR FILING DATE: 1995-03-20
; PRIOR APPLICATION NUMBER: US 08/255,849
; PRIOR FILING DATE: 1994-06-08
; PRIOR APPLICATION NUMBER: US 07/860,710
; PRIOR FILING DATE: 1992-03-30
; PRIOR APPLICATION NUMBER: US 07/523,635
; PRIOR FILING DATE: 1990-05-10
; PRIOR APPLICATION NUMBER: US 07/421,417
; PRIOR FILING DATE: 1989-10-13
; PRIOR APPLICATION NUMBER: US 07/405,370
; PRIOR FILING DATE: 1989-09-11
; PRIOR APPLICATION NUMBER: US 07/403,241
; PRIOR FILING DATE: 1989-09-05
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 17
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Miscellaneous Structure
US-10-252-408-17

Query Match 74.5%; Score 35; DB 15; Length 21;
Best Local Similarity 75.0%; Pred. No. 49;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 STAEGGGG 8
|||:|||||
Db 3 STGDGGG 10

RESULT 13
US-09-864-761-39485
; Sequence 39485, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeonica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 43117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 39485
; LENGTH: 38
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005393.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 5.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 5.4
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.8
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.5
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.1
; OTHER INFORMATION: SWISSPROT HIT: Q01371, EVALUATE 1.30e+00
US-09-864-761-39485

Query Match 74.5%; Score 35; DB 9; Length 38;
Best Local Similarity 75.0%; Pred. No. 85;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 STAEGGGG 8
|||:|||||
Db 12 TTGEGGGG 19

RESULT 14

US-10-101-464A-522
; Sequence 522, Application US/10101464A
; Publication No. US20030046728A1
; GENERAL INFORMATION:
; APPLICANT: Strabala, Timothy
; APPLICANT: Nieuwenhuizen, Nicolaas
; APPLICANT: Higgins, Colleen M.
; TITLE OF INVENTION: Compositions Isolated from Plant Cells
; TITLE OF INVENTION: and Their Use in the Modification of Plant Cell Signaling
; FILE REFERENCE: 11000.1020c2
; CURRENT APPLICATION NUMBER: US/10/101,464A
; CURRENT FILING DATE: 2002-03-18

```
; LENGTH: 157
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: X region
; LOCATION: (33)..(34)
; OTHER INFORMATION: Xaa = any naturally occurring amino acid
; FEATURE:
; NAME/KEY: X region
; LOCATION: (36)..(36)
; OTHER INFORMATION: Xaa = any naturally occurring amino acid
US-10-259-165-344

Query Match      80.9%; Score 38; DB 12; Length 157;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3 AEGGGGR 9
Db      107 AEGGGGR 113

RESULT 6
US-10-346-000A-16
; Sequence 16, Application US/10346000A
; Publication No. US20030215793A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Beatrice H.
; APPLICANT: Shaw, George M.
; APPLICANT: Santiago, Mario L.
; APPLICANT: Cynthia, Rodenburg M.
; APPLICANT: Shadrack, Kamenya
; APPLICANT: Bibollet-Ruche, Frederic
; APPLICANT: Muller, Martin N.
; APPLICANT: Collins, Anthony
; APPLICANT: Wrangham, Richard W.
; APPLICANT: Goodall, Jane
; APPLICANT: Sharp, Paul M.
; TITLE OF INVENTION: Complete Genome Sequence of a Simian Immunodeficiency Virus From
; FILE REFERENCE: P63237US01GP
; CURRENT APPLICATION NUMBER: US/10/346,000A
; CURRENT FILING DATE: 2003-01-16
; PRIOR APPLICATION NUMBER: US 60/349,617
; PRIOR FILING DATE: 2002-01-17
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 16
; LENGTH: 227
; TYPE: PRT
; ORGANISM: Simian immunodeficiency virus
US-10-346-000A-16

Query Match      80.9%; Score 38; DB 12; Length 227;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3 AEGGGGR 9
Db      13 AEGGGGR 19

RESULT 7
US-09-730-525-20
; Sequence 20, Application US/09730525
; Patent No. US20010016956A1
; GENERAL INFORMATION:
; APPLICANT: Ward, Eric
; APPLICANT: Volrath, Sandra
; APPLICANT: Johnson, Marie
; APPLICANT: Potter, Sharon
; TITLE OF INVENTION: Herbicide Tolerant Protox Genes
; Produced by DNA Shuffling
```

```
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. US20010016956Alartis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/09/730,525
; FILING DATE: 05-Dec-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/102,419
; FILING DATE: <Unknown>
; APPLICATION NUMBER: US 09/050,603
; FILING DATE: 30-MAR-1998
; APPLICATION NUMBER: US 09/038,878
; FILING DATE: 11-MAR-1998
; APPLICATION NUMBER: US 08/808,931
; FILING DATE: 28-FEB-1997
; APPLICATION NUMBER: US 60/012,705
; FILING DATE: 28-FEB-1996
; APPLICATION NUMBER: US 60/013,612
; FILING DATE: 28-FEB-1996
; APPLICATION NUMBER: US 60/020,003
; FILING DATE: 21-JUN-1996
; APPLICATION NUMBER: US 08/472,028
; FILING DATE: 06-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: CGC 1847/CIP4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919) 541-8587
; TELEFAX: (919) 541-8689
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 536 amino acids
; TYPE: amino acid
; STRANDEDNESS: No. US20010016956A1 Relevant
; TOPOLOGY: No. US20010016956A1 Relevant
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 20:
US-09-730-525-20

Query Match      78.7%; Score 37; DB 9; Length 536;
Best Local Similarity 87.5%; Pred. No. 4.7e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 STAEGGG 8
Db      43 STIEGGGG 50

RESULT 8
US-09-730-917-20
; Sequence 20, Application US/09730917
; Patent No. US20020073443A1
; GENERAL INFORMATION:
; APPLICANT: Heifetz, Peter
; APPLICANT: Volrath, Sandra
; APPLICANT: Johnson, Marie
; APPLICANT: Ward, Eric
; TITLE OF INVENTION: Herbicide Tolerance Achieved
; THROUGH PLASTID TRANSFORMATION
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 18:21:27; Search time 49.3333 Seconds
(without alignments)
33.929 Million cell updates/sec

Title: US-10-014-658-13
Perfect score: 47 STAEVGGGR 9
Sequence: 1 STAEVGGGR 9

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 684280 seqs, 185983659 residues

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US05_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US05_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10C_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40	85.1	335	12	US-10-306-762-36
2	39	83.0	174	10	US-09-925-300-1200
3	38	80.9	76	12	US-10-029-386-33773
4	38	80.9	157	12	US-10-259-165-10
5	38	80.9	157	12	US-10-259-165-344
6	38	80.9	227	12	US-10-346-000A-16
7	37	78.7	536	9	US-09-730-525-20
8	37	78.7	536	9	US-09-730-917-20
9	36	76.6	121	15	US-10-000-256A-173
10	36	76.6	323	15	US-10-021-811-26
11	36	76.6	726	15	US-10-156-761-12539
12	35	74.5	21	15	US-10-252-408-17
13	35	74.5	38	9	US-09-864-761-39485
14	35	74.5	156	15	US-10-101-464A-522
15	35	74.5	352	12	US-10-259-165-216

16	74.5	1020	15	US-10-205-342-9	Sequence 9, Appli
17	74.5	1061	15	US-10-128-714-3076	Sequence 3076, Ap
18	74.5	1182	15	US-10-128-714-8076	Sequence 8076, Ap
19	74.5	1491	12	US-09-756-247-25	Sequence 25, Appl
20	74.5	1492	12	US-09-981-151A-10	Sequence 10, Appl
21	74.5	1508	12	US-09-756-247-4	Sequence 4, Appli
22	74.5	1508	12	US-10-029-386-28008	Sequence 28008, A
23	72.3	48	9	US-09-864-761-45915	Sequence 45915, A
24	72.3	57	14	US-10-062-254-132	Sequence 132, App
25	72.3	73	15	US-10-059-261-70	Sequence 70, Appl
26	72.3	79	15	US-10-059-261-178	Sequence 178, App
27	72.3	97	15	US-10-074-475-276	Sequence 276, App
28	72.3	130	15	US-10-156-761-11629	Sequence 11629, A
29	72.3	132	15	US-10-156-761-12857	Sequence 12857, A
30	72.3	169	14	US-10-062-254-134	Sequence 134, App
31	72.3	180	12	US-10-029-386-32137	Sequence 32137, A
32	72.3	199	11	US-09-877-843-89	Sequence 89, Appl
33	72.3	199	12	US-09-981-151A-76	Sequence 76, Appl
34	72.3	200	15	US-10-091-438-178	Sequence 178, App
35	72.3	215	15	US-10-156-761-14343	Sequence 14343, A
36	72.3	236	10	US-09-764-864-1047	Sequence 1047, Ap
37	72.3	258	10	US-09-893-666A-2	Sequence 2, Appli
38	72.3	578	12	US-10-029-386-33840	Sequence 33840, A
39	72.3	740	12	US-10-239-663-61	Sequence 61, Appl
40	72.3	2732	12	US-10-238-075-1119	Sequence 1119, Ap
41	72.3	2834	12	US-10-085-959-252	Sequence 252, App
42	70.2	22	12	US-10-029-386-27491	Sequence 27491, A
43	70.2	64	9	US-09-864-761-36560	Sequence 36560, A
44	70.2	80	15	US-10-078-090-181	Sequence 181, App
45	70.2	82	9	US-09-864-761-36140	Sequence 36140, A

ALIGNMENTS

RESULT 1
US-10-306-762-36
; Sequence 36, Application US/10306762
; Publication No. US20030187220A1
; GENERAL INFORMATION:
; APPLICANT: Park, Frances
; APPLICANT: Gajiwala, Ketan S.
; APPLICANT: Buchanan, Sean Grant
; APPLICANT: Sauder, J. Michael
; TITLE OF INVENTION: CRYSTALS AND STRUCTURES OF A FLAVIN
; FILE REFERENCE: MONONUCLEOTIDE BINDING PROTEIN (FMNBP)
; FILE REFERENCE: 52498-20011.00
; CURRENT APPLICATION NUMBER: US/10/306,762
; CURRENT FILING DATE: 2003-04-16
; PRIOR APPLICATION NUMBER: US 60/334,132
; PRIOR FILING DATE: 2001-11-28
; NUMBER OF SEQ ID NOS: 242
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 36
; LENGTH: 335
; TYPE: PRT
; ORGANISM: S. coelicolor A3 2 (21220959)
US-10-306-762-36

Query Match 85.1%; Score 40; DB 12; Length 335;
Best Local Similarity 77.8%; Pred. No. 1.1e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 STAEVGGGR 9
Db 279 ATVEGGGR 287
RESULT 2
US-09-925-300-1200
; Sequence 1200, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:

```
; PRIOR FILING DATE: 1989-09-05
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 17
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Miscellaneous Structure
US-08-406-824A-17

Query Match          74.5%; Score 35; DB 4; Length 21;
Best Local Similarity 75.0%; Pred. No. 12;
Matches 6; Conservative 1; Mismatches 1; Indels 1; Gaps 0;

QY 1 STAEGGGG 8
   ||:||||
Db 3 STGDGGGG 10

RESULT 10
US-09-252-991A-27570
; Sequence 27570, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27570
; LENGTH: 230
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27570

Query Match          74.5%; Score 35; DB 4; Length 230;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 TAEGGGG 8
   ||:||||
Db 67 TADGGGG 73

RESULT 11
US-09-107-532A-4149
; Sequence 4149, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; CORRESPONDENCE ADDRESS:
; ADDRESS: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneka
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 4149:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 523 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...523
; SEQUENCE DESCRIPTION: SEQ ID NO: 4149:
US-09-107-532A-4149

Query Match          74.5%; Score 35; DB 4; Length 523;
Best Local Similarity 87.5%; Pred. No. 2.7e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TAEGGGG 9
   |||||
Db 205 TAEGGTGR 212

RESULT 12
US-09-252-991A-17076
; Sequence 17076, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17076
; LENGTH: 682
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17076

Query Match          74.5%; Score 35; DB 4; Length 682;
Best Local Similarity 85.7%; Pred. No. 3.5e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 AEGGGG 9
   |:||||
Db 396 AOGGGG 402

RESULT 13
US-09-252-991A-25950
; Sequence 25950, Application US/09252991A
; Patent No. 6551795
```


;; CORRESPONDENCE ADDRESS:
;; ADDRESSES: No. 601810sartis Corporation
;; STREET: 520 White Plains Road, P.O. Box 2005
;; CITY: Tarrytown
;; STATE: NY
;; COUNTRY: USA
;; ZIP: 10591-9005
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/808,323
;; FILING DATE: 28-FEB-1996
;; CLASSIFICATION: 800
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 60/012,705
;; FILING DATE: 28-FEB-1996
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 60/013,612
;; FILING DATE: 28-FEB-1996
;; APPLICATION DATA:
;; APPLICATION NUMBER: US 60/020,003
;; FILING DATE: 21-JUN-1996
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Meigs, J. Timothy
;; REGISTRATION NUMBER: 38,241
;; REFERENCE/DOCKET NUMBER: CGC 1846
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (919) 541-8587
;; TELEFAX: (919) 541-8689
;; INFORMATION FOR SEQ ID NO: 20:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 536 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: not relevant
;; TOPOLOGY: not relevant
;; MOLECULE TYPE: protein
US-08-808-323-20

Query Match 78.7%; Score 37; DB 3; Length 536;
Best Local Similarity 87.5%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 STAEGGGG 8
|||
Db 43 STIEGGGG 50

RESULT 6
US-09-050-603A-20
; Sequence 20, Application US/09050603A
; Patent No. 6023012
; GENERAL INFORMATION:
; APPLICANT: Voltath, Sandra
; APPLICANT: Johnson, Marie
; APPLICANT: Potter, Sharon
; APPLICANT: Ward, Eric
; APPLICANT: Heifetz, Peter
; TITLE OF INVENTION: DNA Molecules Encoding Plant
; TITLE OF INVENTION: Protoporphyrogen Oxidase
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6023012artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/102,420B
; FILING DATE: 22-JUN-1998
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/059,164

;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/050,603A
;; FILING DATE: 30-MAR-1998
;; CLASSIFICATION: 800
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/808,931
;; FILING DATE: 28-FEB-1997
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 60/012,705
;; FILING DATE: 28-FEB-1996
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 60/013,612
;; FILING DATE: 28-FEB-1996
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 60/020,003
;; FILING DATE: 21-JUN-1996
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Meigs, J. Timothy
;; REGISTRATION NUMBER: 38,241
;; REFERENCE/DOCKET NUMBER: CGC 1847
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (919) 541-8587
;; TELEFAX: (919) 541-8689
;; INFORMATION FOR SEQ ID NO: 20:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 536 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: not relevant
;; TOPOLOGY: not relevant
;; MOLECULE TYPE: protein
US-09-050-603A-20

Query Match 78.7%; Score 37; DB 3; Length 536;
Best Local Similarity 87.5%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 STAEGGGG 8
|||
Db 43 STIEGGGG 50

RESULT 7
US-09-102-420B-20
; Sequence 20, Application US/09102420B
; Patent No. 6084155
; GENERAL INFORMATION:
; APPLICANT: Voltath, Sandra
; APPLICANT: Johnson, Marie
; APPLICANT: Ward, Eric
; APPLICANT: Heifetz, Peter
; TITLE OF INVENTION: HERBICIDE-TOLERANT PROTOPORPHYRINOGEN
; TITLE OF INVENTION: OXIDASE ("PROTOX")
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6084155artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/102,420B
; FILING DATE: 22-JUN-1998
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/059,164

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 18:21:01 ; Search time 8.22222 Seconds
(without alignments)
46.313 Million cell updates/sec

Title: US-10-014-658-13
Perfect score: 47
Sequence: 1 STAEKGGR 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA: *
1: /cgn2_6/prodata/2/iaa/5A COMB pep: *
2: /cgn2_6/prodata/2/iaa/5B COMB pep: *
3: /cgn2_6/prodata/2/iaa/6A COMB pep: *
4: /cgn2_6/prodata/2/iaa/6B COMB pep: *
5: /cgn2_6/prodata/2/iaa/ECTUS COMB pep: *
6: /cgn2_6/prodata/2/iaa/backfiles1 pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40	85.1	95	3	US-08-905-223-307
2	38	80.9	212	4	US-09-252-991A-23062
3	38	80.9	1116	4	US-09-252-991A-24374
4	37	78.7	536	2	US-08-808-931-20
5	37	78.7	536	3	US-08-808-931-20
6	37	78.7	536	3	US-09-050-603A-20
7	37	78.7	536	3	US-09-102-420B-20
8	37	78.7	536	4	US-09-497-698-20
9	35	74.5	21	4	US-08-406-824A-17
10	35	74.5	230	4	US-09-252-991A-27570
11	35	74.5	523	4	US-09-107-532A-4149
12	35	74.5	682	4	US-09-252-991A-17076
13	34	72.3	144	4	US-09-252-991A-25950
14	34	72.3	161	4	US-09-252-991A-17577
15	34	72.3	267	3	US-08-806-769-145
16	34	72.3	267	3	US-08-906-616-145
17	34	72.3	267	3	US-08-639-075A-145
18	34	72.3	267	3	US-09-004-731-67
19	34	72.3	267	3	US-09-012-431-145
20	34	72.3	267	3	US-08-749-699-67
21	34	72.3	267	3	US-09-012-692-145
22	34	72.3	267	3	US-08-906-613-145
23	34	72.3	267	4	US-09-004-729-67
24	34	72.3	267	4	US-09-252-991A-21390
25	34	72.3	281	1	US-08-040-548-3
26	34	72.3	281	1	US-08-466-344-3
27	34	72.3	339	4	US-09-071-035-80

28	34	72.3	341	4	US-09-328-352-6124
29	34	72.3	361	4	US-09-071-035-78
30	34	72.3	405	4	US-09-198-452A-301
31	34	72.3	407	4	US-09-252-991A-24138
32	34	72.3	441	4	US-09-252-991A-21723
33	34	72.3	496	2	US-08-224-482-2
34	34	72.3	533	1	US-08-040-548-1
35	34	72.3	533	1	US-08-466-344-1
36	34	72.3	533	6	5206152-2
37	34	72.3	668	4	US-09-252-991A-22794
38	34	72.3	718	4	US-09-252-991A-32743
39	34	72.3	738	3	US-08-864-038A-3
40	34	72.3	1384	3	US-08-976-255-11
41	33	70.2	15	2	US-08-572-951-6
42	33	70.2	15	2	US-08-572-951-23
43	33	70.2	174	4	US-09-252-991A-31899
44	33	70.2	246	4	US-09-252-991A-30075
45	33	70.2	252	4	US-09-252-991A-21833

ALIGNMENTS

RESULT 1
US-08-905-223-307
; Sequence 307, Application US/08905223
; Patent No. 622029
; GENERAL INFORMATION:
; APPLICANT: Edwards, Jean-Baptiste D.
; APPLICANT: Duclert, Aymeric
; APPLICANT: Lacroix, Bruno
; TITLE OF INVENTION: 5' ESTs FOR SECRETED PROTEINS
; NUMBER OF SEQUENCES: 503
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 501 West Broadway
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-3505
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Win95
; SOFTWARE: Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/905,223
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Israel, Ned A.
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 307:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 95 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
; ORIGINAL SOURCE:
; ORGANISM: Homo Sapiens
; TISSUE TYPE: Brain
; FEATURE:
; NAME/KEY: sig peptide
; LOCATION: -92--1
; IDENTIFICATION METHOD: Von Heijne matrix
; OTHER INFORMATION: score 6.8
; OTHER INFORMATION: seq ILIASSLPTLSHP/AP

US-08-905-223-307

PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.

PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145313.
PR 27-JUL-1999; 99US-0145318.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 18-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 26-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151920.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.

PR	07-SEP-1999;	99US-0152363.	PD	06-SEP-2000.	
PR	10-SEP-1999;	99US-0153070.	XX	25-FEB-2000;	2000EP-0301439.
PR	13-SEP-1999;	99US-0153758.	XX	25-FEB-1999;	99US-0121825.
PR	15-SEP-1999;	99US-0154018.	PR	05-MAR-1999;	99US-0123180.
PR	16-SEP-1999;	99US-0154039.	PR	09-MAR-1999;	99US-0123548.
PR	20-SEP-1999;	99US-0154779.	PR	23-MAR-1999;	99US-0125788.
PR	22-SEP-1999;	99US-0155139.	PR	25-MAR-1999;	99US-0126264.
PR	23-SEP-1999;	99US-0155486.	PR	29-MAR-1999;	99US-0126785.
PR	24-SEP-1999;	99US-0155659.	PR	01-APR-1999;	99US-0127462.
PR	28-SEP-1999;	99US-0156458.	PR	06-APR-1999;	99US-0128234.
PR	29-SEP-1999;	99US-0156596.	PR	08-APR-1999;	99US-0128714.
PR	04-OCT-1999;	99US-0157117.	PR	16-APR-1999;	99US-0129845.
PR	05-OCT-1999;	99US-0157753.	PR	19-APR-1999;	99US-0130077.
PR	06-OCT-1999;	99US-0157865.	PR	21-APR-1999;	99US-0130449.
PR	07-OCT-1999;	99US-0158029.	PR	23-APR-1999;	99US-0130510.
PR	08-OCT-1999;	99US-0158232.	PR	28-APR-1999;	99US-0130891.
PR	12-OCT-1999;	99US-0158369.	PR	30-APR-1999;	99US-0131449.
PR	13-OCT-1999;	99US-0158293.	PR	30-APR-1999;	99US-0132048.
PR	13-OCT-1999;	99US-0158294.	PR	04-MAY-1999;	99US-0132407.
PR	13-OCT-1999;	99US-0158295.	PR	05-MAY-1999;	99US-0132485.
PR	14-OCT-1999;	99US-0159329.	PR	06-MAY-1999;	99US-0132486.
PR	14-OCT-1999;	99US-0159330.	PR	07-MAY-1999;	99US-0132863.
PR	14-OCT-1999;	99US-0159331.	PR	11-MAY-1999;	99US-0134256.
PR	14-OCT-1999;	99US-0159637.	PR	14-MAY-1999;	99US-0134218.
PR	18-OCT-1999;	99US-0159638.	PR	14-MAY-1999;	99US-0134219.
PR	21-OCT-1999;	99US-0159584.	PR	14-MAY-1999;	99US-0134221.
PR	21-OCT-1999;	99US-0160741.	PR	14-MAY-1999;	99US-0134370.
PR	21-OCT-1999;	99US-0160767.	PR	18-MAY-1999;	99US-0134768.
PR	21-OCT-1999;	99US-0160770.	PR	19-MAY-1999;	99US-0134941.
PR	21-OCT-1999;	99US-0160776.	PR	20-MAY-1999;	99US-0135124.
PR	21-OCT-1999;	99US-0160814.	PR	21-MAY-1999;	99US-0135353.
PR	21-OCT-1999;	99US-0160815.	PR	24-MAY-1999;	99US-0135629.
PR	22-OCT-1999;	99US-0160980.	PR	25-MAY-1999;	99US-0136021.
PR	22-OCT-1999;	99US-0160981.	PR	27-MAY-1999;	99US-0136392.
PR	22-OCT-1999;	99US-0160989.	PR	28-MAY-1999;	99US-0136782.
PR	25-OCT-1999;	99US-0161404.	PR	01-JUN-1999;	99US-0137222.
PR	25-OCT-1999;	99US-0161405.	PR	03-JUN-1999;	99US-0137528.
PR	25-OCT-1999;	99US-0161406.	PR	04-JUN-1999;	99US-0137502.
PR	26-OCT-1999;	99US-0161359.	PR	07-JUN-1999;	99US-0137724.
PR	26-OCT-1999;	99US-0161360.	PR	08-JUN-1999;	99US-0138094.
PR	26-OCT-1999;	99US-0161361.	PR	10-JUN-1999;	99US-0138540.
PR	28-OCT-1999;	99US-0161920.	PR	10-JUN-1999;	99US-0138847.
PR	28-OCT-1999;	99US-0161992.	PR	14-JUN-1999;	99US-0139119.
PR	28-OCT-1999;	99US-0161993.	PR	16-JUN-1999;	99US-0139452.
PR	29-OCT-1999;	99US-0162142.	PR	16-JUN-1999;	99US-0139453.
PR			PR	17-JUN-1999;	99US-0139492.
PR			PR	18-JUN-1999;	99US-0139454.
PR			PR	18-JUN-1999;	99US-0139455.
PR			PR	18-JUN-1999;	99US-0139456.
PR			PR	18-JUN-1999;	99US-0139457.
PR			PR	18-JUN-1999;	99US-0139458.
PR			PR	18-JUN-1999;	99US-0139459.
PR			PR	18-JUN-1999;	99US-0139460.
PR			PR	18-JUN-1999;	99US-0139461.
PR			PR	18-JUN-1999;	99US-0139462.
PR			PR	18-JUN-1999;	99US-0139463.
PR			PR	18-JUN-1999;	99US-0139750.
PR			PR	18-JUN-1999;	99US-0139763.
PR			PR	21-JUN-1999;	99US-0139817.
PR			PR	22-JUN-1999;	99US-0139899.
PR			PR	23-JUN-1999;	99US-0140353.
PR			PR	23-JUN-1999;	99US-0140354.
PR			PR	24-JUN-1999;	99US-0140695.
PR			PR	28-JUN-1999;	99US-0140823.
PR			PR	29-JUN-1999;	99US-0140951.
PR			PR	30-JUN-1999;	99US-0141287.
PR			PR	01-JUL-1999;	99US-0141842.
PR			PR	01-JUL-1999;	99US-0142154.
PR			PR	02-JUL-1999;	99US-0142055.

Query Match 78.7%; Score 37; DB 21; Length 111;
 Best Local Similarity 77.8%; Pred. No. 1.2e+02;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 STRAGGGGR 9
 DB 82 SAASGGGR 90

RESULT 13
 AAGS9597
 ID AAGS9597 standard; Protein; 111 AA.

XX AC AAGS9597;
 XX DT 18-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 77102.

XX KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.

XX OS Arabidopsis thaliana.
 XX PN EP1033405-A2.

```
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 24-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-014287.
PR 01-JUL-1999; 99US-0142842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0143390.
PR 08-JUL-1999; 99US-0143803.
PR 09-JUL-1999; 99US-0143920.
PR 12-JUL-1999; 99US-0143977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.

PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 13-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161320.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 78.7%; Score 37; DB 21; Length 60;
Best Local Similarity 77.8%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qv 1 STRAGGGGR 9
Db 31 SAASGGGGR 39

RESULT 12
AAG11175
ID AAG11175 standard; Protein; 111 AA.
AC AAG11175;
XX
XX 17-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 9793.
XX Protein identification; signal transduction pathway; metabolic pathway;
KW
```

Db 145 AEGGGG 151

RESULT 10
AAG11176
ID AAG11176 standard; Protein; 60 AA.
XX AAG11176;
AC AAG11176;
XX 17-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 9794.
DE Arabidopsis thaliana protein fragment SEQ ID NO: 9794.
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; Genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
XX EP1033405-A2.
PN EP1033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 20-MAY-1999; 99US-0134941.
PR 21-MAY-1999; 99US-0135124.
PR 24-MAY-1999; 99US-0135353.
PR 25-MAY-1999; 99US-0135629.
PR 27-MAY-1999; 99US-0136021.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138947.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143342.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.

CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 106 AA;

Query Match 80.9%; Score 38; DB 22; Length 106;

Best Local Similarity 77.8%; Pred. No. 78;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 STAEGGGGR 9

Db 72 STVRGGGR 80

RESULT 8

AA084729

ID AA084729 standard; Protein; 124 AA.

XX AC AA084729;

XX DT 07-NOV-2001 (first entry)

XX DE Human immune/haematopoietic antigen SEQ ID NO:12322.

XX KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;

XX KW Cytostatic; Gene therapy; vaccine; metastasis.

XX OS Homo sapiens.

XX PN WC020151782-A2.

XX PD 09-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US01354.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0183350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 20-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.

PR 14-AUG-2000; 2000US-0225759.

PR 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226681.

PR 22-AUG-2000; 2000US-0226868.

PR 22-AUG-2000; 2000US-0227182.

PR 23-AUG-2000; 2000US-0227009.

PR 30-AUG-2000; 2000US-0228924.

PR 01-SEP-2000; 2000US-0229287.

PR 01-SEP-2000; 2000US-0229343.

PR 01-SEP-2000; 2000US-0229344.

PR 01-SEP-2000; 2000US-0229345.

PR 05-SEP-2000; 2000US-0229509.

PR 05-SEP-2000; 2000US-0229513.

PR 06-SEP-2000; 2000US-0230437.

PR 06-SEP-2000; 2000US-0230438.

PR 08-SEP-2000; 2000US-0231242.

PR 08-SEP-2000; 2000US-0231243.

PR 08-SEP-2000; 2000US-0231244.

PR 08-SEP-2000; 2000US-0231413.

PR 08-SEP-2000; 2000US-0231414.

PR 08-SEP-2000; 2000US-0232080.

PR 08-SEP-2000; 2000US-0232081.

PR 12-SEP-2000; 2000US-0231968.

PR 14-SEP-2000; 2000US-0232397.

PR 14-SEP-2000; 2000US-0232398.

PR 14-SEP-2000; 2000US-0232399.

PR 14-SEP-2000; 2000US-0232400.

PR 14-SEP-2000; 2000US-0232401.

PR 14-SEP-2000; 2000US-0233063.

PR 14-SEP-2000; 2000US-0233064.

PR 14-SEP-2000; 2000US-0233065.

PR 21-SEP-2000; 2000US-0234223.

PR 21-SEP-2000; 2000US-0234274.

PR 25-SEP-2000; 2000US-0234597.

PR 25-SEP-2000; 2000US-0234998.

PR 26-SEP-2000; 2000US-0235484.

PR 27-SEP-2000; 2000US-0235834.

PR 27-SEP-2000; 2000US-0235836.

PR 29-SEP-2000; 2000US-0236327.

PR 29-SEP-2000; 2000US-0236367.

PR 29-SEP-2000; 2000US-0236368.

PR 29-SEP-2000; 2000US-0236369.

PR 29-SEP-2000; 2000US-0236370.

PR 02-OCT-2000; 2000US-0236802.

PR 02-OCT-2000; 2000US-0237037.

PR 02-OCT-2000; 2000US-0237038.

PR 02-OCT-2000; 2000US-0237039.

PR 02-OCT-2000; 2000US-0237040.

PR 13-OCT-2000; 2000US-0239935.

PR 13-OCT-2000; 2000US-0239937.

PR 20-OCT-2000; 2000US-0240960.

PR 20-OCT-2000; 2000US-0241221.

PR 20-OCT-2000; 2000US-0241785.

PR 20-OCT-2000; 2000US-0241786.

PR 20-OCT-2000; 2000US-0241787.

PR 20-OCT-2000; 2000US-0241808.

PR 20-OCT-2000; 2000US-0241809.

PR 01-NOV-2000; 2000US-0241826.

PR 01-NOV-2000; 2000US-0241826.

PR 08-NOV-2000; 2000US-0246474.

PR 08-NOV-2000; 2000US-0246475.

PR 08-NOV-2000; 2000US-0246476.

PR 08-NOV-2000; 2000US-0246477.

PR 08-NOV-2000; 2000US-0246478.

PR 08-NOV-2000; 2000US-0246523.

PR 08-NOV-2000; 2000US-0246523.

PR 08-NOV-2000; 2000US-0246524.

PR 08-NOV-2000; 2000US-0246525.

PR 08-NOV-2000; 2000US-0246526.

PR 08-NOV-2000; 2000US-0246527.

PR 08-NOV-2000; 2000US-0246528.

PR 08-NOV-2000; 2000US-0246532.

PR 08-NOV-2000; 2000US-0246609.

PT which may have cytokine, immune, regulatory, haematopoiesis
PT regulating, anti-inflammatory or tumour inhibition activity
XX
PS Claim 34; Page 457; 577pp; English.
XX
CC AAX51787 to AAX52019 represent 5' expressed sequence tags (ESTs) for
CC human secreted proteins, and encode the proteins given in AAY12987 to
CC AAY13219, respectively. The proteins given represent the signal peptide
CC and an N-terminal fragment of a secreted protein. The nucleic acid
CC sequences can be used for producing secreted human gene products. They
CC can also be used to develop products for diagnosis and therapy. The
CC proteins obtained may have cytokine activity, cell
CC proliferation/differentiation activity, haematopoiesis regulating
CC activity, tissue growth regulating activity, reproductive hormone
CC regulating activity, chemotactic/chemokinetic activity, haemostatic and
CC thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, tumour inhibition activity or other activities. The products
CC can be used in forensic, gene therapy and chromosome mapping procedures.
CC The sequences can also be used for obtaining corresponding promoter
CC sequences. The nucleic acids encoding the signal peptide can be used for
CC directing extracellular secretion of a polypeptide or the insertion of a
CC polypeptide into a membrane, or importing a polypeptide into a cell.
XX
SQ Sequence 95 AA;

Query Match 85.1%; Score 40; DB 20; Length 95;
Best Local Similarity 88.9%; Pred. No. 34;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 STAEGGGGG 9
DB 21 SHAEGGGG 29

RESULT 4
ID AAO02150 standard; Protein; 132 AA.
AC AAO02150;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 16042.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200164835-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US04927.
XX
PR 28-FEB-2000; 2000US-0515126.
PR 18-MAY-2000; 2000US-0577409.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-514838/56.
DR N-PSDB; AAI82081.
XX
PT Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX
PS Claim 20; SEQ ID NO 16042; 1399pp + Sequence Listing; English.
XX

CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAQ00010-AAQ13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 132 AA;

Query Match 83.0%; Score 39; DB 22; Length 132;
Best Local Similarity 87.5%; Pred. No. 67;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 STAEGGGG 8
DB 44 STAQGGG 51

RESULT 5
ID AAB56622 standard; Protein; 174 AA.
AC AAB56622;
XX
DT 13-MAR-2001 (first entry)
XX
DE Human prostate cancer antigen protein sequence SEQ ID NO:1200.
XX
KW Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
KW neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;
KW vulnary; gastrointestinal; nephrotropic; antiinfective; gynaecological;
KW antibacterial; gene therapy; neural; immune; reproductive; renal;
KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KW wound; infectious disease.
XX
OS Homo sapiens.
XX
PN WO200055174-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05988.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2000-587513/55.
DR N-PSDB; AAI15825.
XX
PT Prostate cancer associated gene sequences, referred to as prostate
PT cancer antigens, useful for treatment, prevention, and diagnosis of
PT disorders such as prostate cancer -
XX
PS Claim 11; Page 1612; 2338pp; English.
XX
CC AAI15866 to AAI16505 encode the human prostate cancer associated
CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
CC The prostate cancer antigens can have neuroprotective, cyostatic,
CC cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,
CC nephrotropic, antiinfective, gynaecological and antibacterial activities,
CC and can be used in gene therapy. The prostate cancer antigen

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 17:22:25, Search time 25 Seconds
(without alignments)
57.142 Million cell updates/sec

Title: US-10-014-658-13

Perfect score: 47

Sequence: 1 STAREGGGR 9

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: A_Geneseq_13Jun03.*

```
1: /SIDSL/gcgdata/geneseq/geneseq-embl/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq-embl/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq-embl/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq-embl/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq-embl/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq-embl/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq-embl/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq-embl/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq-embl/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq-embl/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq-embl/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq-embl/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq-embl/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq-embl/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq-embl/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq-embl/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq-embl/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq-embl/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq-embl/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq-embl/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq-embl/AA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq-embl/AA2001.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseq-embl/AA2002.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseq-embl/AA2003.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	47	100.0	9	21	AA1980
2	43	91.5	53	22	AA1981
3	40	85.1	95	20	AA1982
4	39	83.0	132	22	AA1983
5	39	83.0	174	21	AA1984
6	39	83.0	1605	23	AA1985
7	38	80.9	106	22	AA1986
8	38	80.9	124	22	AA1987
9	38	80.9	384	22	AA1988

10	37	78.7	60	21	AA1989	Arabidopsis thalia
11	37	78.7	60	21	AA1990	Arabidopsis thalia
12	37	78.7	111	21	AA1991	Arabidopsis thalia
13	37	78.7	111	21	AA1992	Arabidopsis thalia
14	37	78.7	120	21	AA1993	Arabidopsis thalia
15	37	78.7	536	18	AA1994	Rape protoplasm
16	37	78.7	536	18	AA1995	Rape protoplasm
17	37	78.7	536	22	AA1996	Oilseed rape prot
18	37	78.7	536	22	AA1997	Oilseed rape prot
19	37	78.7	536	22	AA1998	Oilseed rape prot
20	37	78.7	536	22	AA1999	Oilseed rape prot
21	36	76.6	88	22	AA2000	Human polypeptide
22	36	76.6	91	21	AA2001	Human polypeptide
23	36	76.6	121	23	AA2002	Human secreted pro
24	36	76.6	174	22	AA2003	Drosophila melanog
25	36	76.6	328	22	AA2004	Novel human diagno
26	36	76.6	449	22	AA2005	Drosophila melanog
27	36	76.6	741	20	AA2006	N. meningitidis st
28	35	74.5	38	22	AA2007	Human liver peptid
29	35	74.5	38	22	AA2008	Human liver peptid
30	35	74.5	38	22	AA2009	Protein #6186 enco
31	35	74.5	38	22	AA2010	Human brain expres
32	35	74.5	38	22	AA2011	Human bone marrow
33	35	74.5	38	22	AA2012	Peptide #6122 enco
34	35	74.5	38	22	AA2013	Peptide #6966 enco
35	35	74.5	38	23	AA2014	Human polypeptide
36	35	74.5	41	22	AA2015	Human polypeptide
37	35	74.5	53	22	AA2016	Human polypeptide
38	35	74.5	70	22	AA2017	Human immune/haema
39	35	74.5	74	21	AA2018	Human secreted pro
40	35	74.5	85	22	AA2019	Human polypeptide
41	35	74.5	111	22	AA2020	Human polypeptide
42	35	74.5	140	21	AA2021	Arabidopsis thalia
43	35	74.5	150	21	AA2022	Arabidopsis thalia
44	35	74.5	156	21	AA2023	Eucalyptus grandis
45	35	74.5	183	22	AA2024	Drosophila melanog

ALIGNMENTS

RESULT 1

AA1989

ID AA1989 standard; peptide; 9 AA.

XX AA1989

AC AA1989

XX 27-MAR-2000 (first entry)

DT Human antithrombin III variant Aa (residues 385-393).

XX Human; antithrombin III; ATIII variant Aa; elastase-resistant;

XX Igg activated neutrophil resistant; anti-thrombin activity; heparin;

XX anti-factor Xa activity; blood clotting disorder; sepsis; trauma; stroke;

XX thrombin activation-related pathological symptom; restenosis; thrombosis;

XX acute respiratory distress syndrome; thromboembolism; reocclusion.

XX Homo sapiens.

OS Synthetic.

XX Key

FT Misc-difference 4 Location/Qualifiers

FT Misc-difference 5 /note= "ATIII.N135A Val at 388 is substituted by Glu"

FT Misc-difference 6 /note= "ATIII.N135A Val at 389 is substituted by Gly"

FT Misc-difference 7 /note= "ATIII.N135A Ile at 390 is substituted by Gly"

FT Misc-difference 8 /note= "ATIII.N135A Ala at 391 is substituted by Gly"

XX WO9558098-A2.

XX 18-NOV-1999.

PD

RL Science 287:2185-2195(2000).
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
DR EMBL; AE003765; AAF56794.2; -.
DR EMBL; AE003765; AAF56795.2; -.
DR HSP; P12111; 2KNT.
DR FlyBase; FBgn003137; Pdn.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG_c2.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR002223; Kunitz_BPTI.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR002221; WAP.
DR Pfam; PF00047; ig; 3.
DR Pfam; PF00014; Kunitz_BPTI; 12.
DR Pfam; PF00090; tsp_1; 5.
DR Pfam; PF00095; wap; 1.
DR PRINTS; PR00759; BASICPTASE.
DR ProDom; PD000222; Kunitz_BPTI; 12.
DR SMART; SM00408; IGC2; 3.
DR SMART; SM00131; KU; 12.
DR SMART; SM00209; TSP1; 7.
DR SMART; SM00217; WAP; 1.
DR PROSITE; PS00317; 4_DISULFIDE_CORE; 1.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 11.
DR PROSITE; PS00279; BPTI_KUNITZ_2; 12.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS00835; IG-like; 3.
DR PROSITE; PS00932; TSP1; 5.
KW Alternative splicing; immunoglobulin domain; Protease inhibitor;
KW Serine protease inhibitor.
FT VARSPPLIC 2803 2803 L -> SVPEV (IN SHORT ISOFORM).
FT VARSPPLIC 2844 2854 FNKTMEDSGI -> VASPLHNAV (IN SHORT
ISOFORM).
FT VARSPPLIC 2855 3060 MISSING (IN SHORT ISOFORM).
SQ SEQUENCE 3060 AA; 331579 MW; ACA31D3E5558C7C0 CRC64;
Query Match 75.0%; Score 33; DB 5; Length 3060;
Best Local Similarity 55.6%; Pred. No. 2.3e+03;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
QY 1 STEVEGAGR 9
Db 1309 TTEIGSGQ 1317

Search completed: December 11, 2003, 18:28:26
Job time : 24.3333 secs

```
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=2202145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach P.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Canarotte G., Cannavan F., Cardoso J., Chambergo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463(2002).
DR EMBL; AE011713; AAM35706.1; -.
DR InterPro; IPR002173; PFKB
DR InterPro; IPR000719; Prot_Kinase.
DR InterPro; IPR002139; Ribokinase.
DR Pfam; PF00294; pfkB; 1.
DR PRINTS; PR00990; RIBOKINASE.
DR PROSITE; PS00583; PFKB_KINASES_1; 1.
DR PROSITE; PS00584; PFKB_KINASES_2; 1.
DR PROSITE; PS0107; PROTEIN_KINASE_ATP; 1.
KW Kinase; Complete proteome.
SQ SEQUENCE 316 AA; 32113 MW; 20059AE5AB870ECP CRC64;

Query Match 75.0%; Score 33; DB 16; Length 316;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 STEVEGA 7
DB 299 STEVEGA 305

RESULT 11
QSPCHO PRELIMINARY; PRT; 318 AA.
AC QSPCHO;
DT 01-OCT-2002 (TReMBLrel. 22, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Ribokinase.
GN RBSK OR XCC0764.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=2202145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach P.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Canarotte G., Cannavan F., Cardoso J., Chambergo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463(2002).
DR EMBL; AE011713; AAM35706.1; -.
DR InterPro; IPR002173; PFKB
DR InterPro; IPR000719; Prot_Kinase.
DR InterPro; IPR002139; Ribokinase.
DR Pfam; PF00294; pfkB; 1.
DR PRINTS; PR00990; RIBOKINASE.
DR PROSITE; PS00583; PFKB_KINASES_1; 1.
DR PROSITE; PS00584; PFKB_KINASES_2; 1.
DR PROSITE; PS0107; PROTEIN_KINASE_ATP; 1.
KW Kinase; Complete proteome.
SQ SEQUENCE 316 AA; 32113 MW; 20059AE5AB870ECP CRC64;

Query Match 75.0%; Score 33; DB 16; Length 316;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 STEVEGA 7
DB 299 STEVEGA 305

RESULT 11
QSPCHO PRELIMINARY; PRT; 318 AA.
AC QSPCHO;
DT 01-OCT-2002 (TReMBLrel. 22, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Ribokinase.
GN RBSK OR XCC0764.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=2202145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach P.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Canarotte G., Cannavan F., Cardoso J., Chambergo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463(2002).
DR EMBL; AE011713; AAM35706.1; -.
DR InterPro; IPR002173; PFKB
DR InterPro; IPR000719; Prot_Kinase.
DR InterPro; IPR002139; Ribokinase.
DR Pfam; PF00294; pfkB; 1.
DR PRINTS; PR00990; RIBOKINASE.
DR PROSITE; PS00583; PFKB_KINASES_1; 1.
DR PROSITE; PS00584; PFKB_KINASES_2; 1.
DR PROSITE; PS0107; PROTEIN_KINASE_ATP; 1.
KW Kinase; Complete proteome.
SQ SEQUENCE 316 AA; 32113 MW; 20059AE5AB870ECP CRC64;

Query Match 75.0%; Score 33; DB 16; Length 318;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 STEVEGA 7
DB 299 STEVEGA 305

RESULT 12
Q41113 PRELIMINARY; PRT; 339 AA.
ID Q41113;
AC Q41113;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE bZIP transcriptional repressor ROM1.
OS Phaseolus vulgaris (Kidney bean) (French bean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Phaseolus.
OX NCBI_TaxID=3885;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96314865; PubMed=8758983;
RA Chern M.S., Eiben H.G., Bustos M.M.;
RT "The developmentally regulated bZIP factor ROM1 modulates
RT transcription from lectin and storage protein genes in bean embryos.";
RL Plant J. 10:135-148(1996).
CC -!- SIMILARITY; BELONGS TO THE BZIP FAMILY.
DR EMBL; U57389; AAB36514.1; -.
DR HSSP; P03069; IYSA.
DR TRANSFAC; T02809; -.
DR InterPro; IPR001630; Leuzip_CREB.
DR InterPro; IPR004827; TF_bZIP.
DR Pfam; PF00170; bzip; 1.
DR PRINTS; PR00041; LEUZIPPRCREB.
DR SMART; SM00338; BRLZ; 1.
DR PROSITE; PS00036; BZIP_BASIC; 1.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 339 AA; 35883 MW; 10DE71E961D805FB CRC64;

Query Match 75.0%; Score 33; DB 10; Length 339;
Best Local Similarity 75.0%; Pred. No. 2.1e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 STEVEGAG 8
DB 104 STEIEGK 111

RESULT 13
Q8FNB5 PRELIMINARY; PRT; 345 AA.
ID Q8FNB5
AC Q8FNB5;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
```

```
DR EMBL; AEO12549; AAM43437.1; -.
DR InterPro; IPR005887; alpha man.
DR TIGRFAMs; TIGR01180; anan2_put; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 824 AA; 89711 MW; 863CE00B3D9B2A2 CRC64;

Query Match 81.8%; Score 36; DB 16; Length 824;
Best Local Similarity 77.8%; Pred. No. 1.3e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 STEVEGAGR 9
Db 331 SVQVEGAGR 339

RESULT 5
QNS07 ID QNS07 PRELIMINARY; PRT; 135 AA.
AC QNS07;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE Large-conductance mechanosensitive channel.
GN CGL0879.
OS Corynebacterium glutamicum (Brevibacterium flavum).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=17118;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RA Nakagawa S.;
RL "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
DR EMBL; AP005276; BAB98272.1; -.
DR InterPro; IPR001185; MS channel.
DR Pfam; PF01741; MscL; 1.
DR PRINTS; PR01264; MECHCHANNEL.
DR ProDom; PD007253; MS channel; 1.
DR TIGRFAMs; TIGR00220; mscL; 1.
KW Complete proteome.
SQ SEQUENCE 135 AA; 14535 MW; 9D69CD0921D1E594 CRC64;

Query Match 77.3%; Score 34; DB 16; Length 135;
Best Local Similarity 87.5%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 STEVEGAG 8
Db 48 STEVEGLG 55

RESULT 6
QNS07 ID QNS07 PRELIMINARY; PRT; 157 AA.
AC QNS07;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE Hypothetical protein CPE0594.
GN CPE0594.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / Type A;
RX PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Hoshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
flesh-eater.";
```

```
Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR EMBL; AP00187; BAB80300.1; -.
DR InterPro; IPR000096; NUDIX_hydrolase.
DR Pfam; PF00293; NUDIX; 1.
DR PROSITE; PS00893; NUDIX; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 157 AA; 17812 MW; E6FD837CD596890 CRC64;

Query Match 75.0%; Score 33; DB 16; Length 157;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 STEVEGA 7
Db 105 STEVEGA 111

RESULT 7
Q9KZ62 ID Q9KZ62 PRELIMINARY; PRT; 259 AA.
AC Q9KZ62;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE Putative lipoprotein.
DE Putative lipoprotein.
GN SCO3086 OR SC825.27.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomyces; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21596410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieffer H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieffer T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch B., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL939115; CAB89458.1; -.
KW Lipoprotein; Complete proteome.
SQ SEQUENCE 259 AA; 26022 MW; 1BD1B498FCF3EB5C CRC64;

Query Match 75.0%; Score 33; DB 16; Length 259;
Best Local Similarity 62.5%; Pred. No. 1.6e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 TEVEGAGR 9
Db 110 TDIEGAGR 117

RESULT 8
Q9XAL6 ID Q9XAL6 PRELIMINARY; PRT; 263 AA.
AC Q9XAL6;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DE Putative hydroxylase.
GN SCO6714 OR SC4C6.24C.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomyces; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
```

```

DT 01-JUN-1998 (TRENBLrel. 06, Created)
DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE 2,4-dienoyl-CoA reductase.
GN FADH OR RV1175C OR MT1212 OR MTV005.11C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterinase; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Broch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigleier K., Gas S., Barry C.E. III, Tekala P.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feldwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544 (1998).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RX Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RA "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL010186; CAA15852.1; -.
DR EMBL; AE006998; AAK45469.1; -.
DR TIGR; MT1212; -.
DR Tuberculast; RV1175c; -.
DR InterPro; IPR000759; Adrnkx_reductase.
DR InterPro; IPR001064; Crystallin.
DR InterPro; IPR001327; FAD pyr redox.
DR InterPro; IPR000205; NAD binding.
DR InterPro; IPR001155; Oxidored FMN.
DR InterPro; IPR001100; Pyr redox.
DR Pfam; PF00724; oxidored_FMN; 1.
DR Pfam; PF00070; Pyr_redox; 1.
DR PRINTS; PR00419; ADXRDPTASE.
DR PRINTS; PR00368; FADPNR.
DR PRINTS; PR00411; PNDRTDPTASEI.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 1.
KW Complete proteome.
FT CONFLICT 210
FT CONFLICT 210
SQ SEQUENCE 674 AA; 72901 MW; 874787F74ED3EE0A5 CRC64;

Query Match 81.8%; Score 36; DB 16; Length 674;
Best Local Similarity 87.5%; Pred. No. 1.1e+02;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 STEVEGAG 8
Db 237 ATEVEGAG 244

RESULT 3
Q8TH00 PRELIMINARY; PRT; 711 AA.
AC Q8TH00;
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Predicted superfamily II helicase.
GN MK0111.

```

```

OS Methanopyrus kandleri.
OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
OC Methanopyrus.
OX NCBI_TaxID=2320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AV19 / DSM 6324 / JCM 9639;
RX MEDLINE=21927647; PubMed=11930014;
RA Slesarev A.I., Mezhevaya K.V., Makarova K.S., Polushin N.N.,
RA Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,
RA Natale D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,
RA Malykh A.G., Koonin E.V., Kozhavkin S.A.;
RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19
and monophyly of archaeal methanogens.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649 (2002).
DR EMBL; AE010310; AAM01328.1; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR000445; HHH.
DR InterPro; IPR003583; HHH_1.
DR Pfam; PF00270; DEAD; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR Pfam; PF00633; HHH; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELICC; 1.
DR SMART; SM00278; HHH1; 2.
DR Helicase; Complete proteome.
KW SEQUENCE 711 AA; 79588 MW; 32EBE5848BE37E4E CRC64;

Query Match 81.8%; Score 36; DB 17; Length 711;
Best Local Similarity 87.5%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TEVEGAGR 9
Db 655 TEVEGVGR 662

RESULT 4
Q8P357 PRELIMINARY; PRT; 824 AA.
AC Q8P357;
ID Q8P357;
DT 01-OCT-2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Hypothetical protein XCC4221.
DE XCC4221.
GN Xanthomonas campestris (pv. campestris)
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=2022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.B.A.,
RA Camarotte G., Canavan P., Cardoso J., Chambergo F., Cispina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Fornighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locall E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities.";
RL Nature 417:459-463 (2002).

```

DR GO; GO:0008270; F:zinc ion binding activity; IDA.
DR InterPro; IPR004457; ZnF_ZPR1.

DR Pfam; PF03367; ZPR1; 2.

DR SMART; SM00709; ZPR1; 2.

DR TIGRFAMs; TIGR00310; ZPR1_znf; 2.

KW Nuclear Protein; Zinc-finger. C4-TYPE.

FT ZN_FING 51 83

FT ZN_FING 259 291

SQ SEQUENCE 459 AA; 50715 MW; 771D38C7B806044F CRC64;

Query Match 70.5%; Score 31; DB 1; Length 459;

Best Local Similarity 55.6%; Pred No. 75;

Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 STEVEGAGR 9

DB 87 NTEIQSAGR 95

RESULT 15

TKTA_CRAPL

ID TKTA_CRAPL STANDARD; PRT; 679 AA.

AC Q42675;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DE 01-NOV-1997 (Rel. 35, Last annotation update)

DE Transketolase 10 (EC 2.2.1.1) (TK).

GN TK10.

OS Craterostigma plantagineum.

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

OC Asteridae; lamids; Lamiales; Lamiales incertae sedis; Torenieae;

OC Craterostigma.

OC NCBI_TaxID=4153;

OX NCBI_TaxID=4153;

RN [1]

SEQUENCE FROM N.A.

RX MEDLINE=95163594; PubMed=7859749;

RA Bernacchia G., Schwall G., Lottspeich F., Salamini F., Bartels D.;

RT "The transketolase gene family of the resurrection plant

Craterostigma plantagineum: differential expression during the

rehydration phase";

RL EMBO J. 14:610-618(1995).

CC -!- FUNCTION: COULD BE INVOLVED IN THE CONVERSION OF SUGARS, WHICH ARE

A MAJOR PHENOMENON IN THE REHYDRATION PROCESS.

CC -!- CATALYTIC ACTIVITY: Sedoheptulose 7-phosphate + D-glyceraldehyde

3-phosphate = D-ribose 5-phosphate + D-xylulose 5-phosphate.

CC -!- COFACTOR: THIAMINE PYROPHOSPHATE.

CC -!- TISSUE SPECIFICITY: Leaves.

CC -!- INDUCTION: By rehydration.

CC -!- SIMILARITY: BELONGS TO THE TRANSKETOLASE FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

CC -----

CC EMBL; Z46647; CAA86608.1; -

DR PIR; S54299; S54299.

DR HSSP; P23254; 1TRK.

DR InterPro; IPR005478; BacTransketolase.

DR InterPro; IPR005476; Transketolase_C.

DR InterPro; IPR005475; Transketolase_CR.

DR InterPro; IPR005474; Transketolase_N.

DR Pfam; PF02779; transket_pyr; 1.

DR Pfam; PF00456; transketolase; 1.

DR Pfam; PF02780; transketolase_C; 1.

DR TIGRFAMs; TIGR00232; tktlase_bact; 1.

DR PROSITE; PS00801; TRANSKETOLASE_1; 1.

DR PROSITE; PS00802; TRANSKETOLASE_2; 1.

KW Transferase; Thiamine pyrophosphate; Multigene family.

SQ SEQUENCE 679 AA; 73130 MW; 1109092E136A345B CRC64;

Query Match

Best Local Similarity 70.5%; Score 31; DB 1; Length 679;

Matches 6; Conservative 0; Mismatches 2; Indels 0;

Gaps 0;

QY 2 TEVEGAGR 9

DB 544 TSVEGVGR 551

Search completed: December 11, 2003, 18:21:03

Job time : 6.66667 secs


```

CC N-TERMINUS.
CC -!- SIMILARITY: BELONGS TO THE DELTA ENDOTOXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L32019; AAA22341.1; -.
CC HSP: P02965; 1C1Y.
CC InterPro: IPR001178; Endotoxin.
CC InterPro: IPR005638; endotoxin_C.
CC InterPro: IPR005639; endotoxin_N.
CC Pfam: PF00555; endotoxin_1.
CC Pfam: PF03944; endotoxin_C; 1.
CC Pfam: PF03945; endotoxin_N; 1.
CC Toxin; Sporulation.
CC SQ SEQUENCE 1167 AA; 132760 MW; 3B7357D14E655FC7 CRC64;
CC
CC Query Match 72.7%; Score 32; DB 1; Length 1167;
CC Best Local Similarity 75.0%; Pred. No. 1.2e+02;
CC Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
CC
CC QY 2 TEVEGAGR 9
CC ||:|:|
CC Db 97 TELEGLGR 104
CC
CC RESULT 9
CC SECE_ECOLI STANDARD; PRT; 127 AA.
CC AC PL6920;
CC DT 01-AUG-1990 (Rel. 15, Created)
CC DT 01-AUG-1990 (Rel. 15, Last sequence update)
CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
CC DE Preprotein translocase secE subunit.
CC GN SECE OR PRLG OR B3981 OR C4936 OR Z5554 OR ECS4904.
CC OS Escherichia coli,
CC OS Escherichia coli O6, and
CC OS Escherichia coli O157:H7.
CC OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
CC OC Enterobacteriaceae; Escherichia.
CC OX NCBI_TaxID=562, 217592, 83334;
CC [1]
CC RN SEQUENCE FROM N.A.
CC RP MEDLINE=90170882; PubMed=2137819;
CC RA Downing W.L., Sullivan S.L., Gottesman M.E., Dennis P.P.;
CC RT "Sequence and transcriptional pattern of the essential Escherichia
CC coli secE-nusG operon.";
CC RL J. Bacteriol. 172:1621-1627(1990).
CC [2]
CC RN SEQUENCE FROM N.A.
CC RP MEDLINE=9378734; PubMed=2673920;
CC RA Schatz P.J., Riggs P.D., Jacq A., Fath M.J., Beckwith J.;
CC RT "The secE gene encodes an integral membrane protein required for
CC protein export in Escherichia coli.";
CC RL Genes Dev. 3:1035-1044(1989).
CC [3]
CC RN SEQUENCE FROM N.A.
CC RP STRAIN=K12 / MG1655;
CC RX MEDLINE=94089392; PubMed=8265357;
CC RA Blattner F.R., Burland V.D., Plunkett G. III, Sofia H.J.,
CC RA Daniels D.L.;
CC RT "Analysis of the Escherichia coli genome. IV. DNA sequence of the
CC region from 89.2 to 92.8 minutes.";
CC RL Nucleic Acids Res. 21:5408-5417(1993).
CC [4]
CC RN SEQUENCE FROM N.A.
CC RP STRAIN=O6:H1 / CFT073 / ATCC 700928;
CC RX MEDLINE=22388234; PubMed=12471157;

```

```

RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Raeko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Ferna N.T.;
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RA "Extensive mosaic structure revealed by the complete genome sequence
RA of uropathogenic Escherichia coli.";
RA Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
RA [5]
RA SEQUENCE FROM N.A.
RA STRAIN=O157:H7 / EDL933 / ATCC 700927;
RA MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Goddard E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RA "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RA Nature 409:529-533(2001).
RA [6]
RA SEQUENCE FROM N.A.
RA STRAIN=O157:H7 / RIMD 0509952;
RA MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobo T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RA "Complete genome sequence of enterohemorrhagic Escherichia coli
RA O157:H7 and genomic comparison with a laboratory strain K-12.";
RA DNA Res. 8:11-22(2001).
RA [7]
RA TOPOLOGY.
RA MEDLINE=91266903; PubMed=2050112;
RA Schatz P.J., Bieker K.L., Ottemann K.M., Silhavy T.J., Beckwith J.;
RA "One of three transmembrane stretches is sufficient for the
RA functioning of the SecE protein, a membrane component of the E. coli
RA secretion machinery.";
RA EMBL J. 10:1749-1757(1991).
RA [8]
RA SUBUNITS.
RA MEDLINE=20402345; PubMed=10944122;
RA Yahr T.L., Wickner W.T.;
RA "Evaluating the oligomeric state of SecYEG in preprotein
RA translocase.";
RA EMBL J. 19:4393-4401(2000).
RA [9]
RA SUBUNITS.
RA MEDLINE=20164453; PubMed=10699927;
RA Manting E.H., van Der Does C., Remigy H., Engel A., Driessen A.J.;
RA "SecYEG assembles into a tetramer to form the active protein
RA translocation channel.";
RA EMBL J. 19:852-861(2000).
RA [10]
RA MUTANTS.
RA MEDLINE=95196752; PubMed=7889938;
RA Flower A.M., Osborne R.S., Silhavy T.J.;
RA "The allele-specific synthetic lethality of prlA-prlG double mutants
RA predicts interactive domains of SecY and SecE.";
RA EMBL J. 14:884-893(1995).
RA CC -!- FUNCTION: ESSENTIAL FOR PROTEIN EXPORT.
RA CC -!- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLLOCATION APPARATUS
RA CC WHICH COMPRISE SECA, SECB, SECD, SECE, SECF, SECG AND SECY. THE
RA CC TRANSLLOCATION CHANNEL SEEMS TO BE COMPOSED OF A SECA HOMODIMER AND
RA CC FOUR COPIES OF A SECYEG COMPLEX.
RA CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
RA CC -!- SIMILARITY: Belongs to the secE/SEC61-gamma family.
RA CC -----
RA This SWISS-PROT entry is copyrighted. It is produced through a collaboration
RA between the Swiss Institute of Bioinformatics and the EMBL outstation -
RA the European Bioinformatics Institute. There are no restrictions on its
RA use by non-profit institutions as long as its content is in no way
RA modified and this statement is not removed. Usage by and for commercial
RA entities requires a license agreement (See http://www.isb-sib.ch/announce/

```



```
Query Match 72.7%; Score 32; DB 1; Length 324;
Best Local Similarity 66.7%; Pred. No. 33;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 STEVEGAGR 9
DB 18 SWEIRGAGR 26

RESULT 4
RL4_URECA
ID RL4_URECA STANDARD; PRT; 386 AA.
AC P49165;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 60S ribosomal protein L4 (L1).
GN RPL4 OR RPL1.
OS Urechis caupo (Innkeeper worm) (Spoonworm).
OC Eukaryota; Metazoa; Echiura; Xenopneusta; Urechidae; Urechis.
OX NCBI_TaxID=6431;
RN [1]
RP SEQUENCE FROM N.A.
RA Rosenthal E.T.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE L4E FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; U30495; AAA74021.1; -.
DR PIR; T12048; T12048.
DR InterPro; IPR002136; Ribosomal L4/L1E.
DR Pfam; PF00573; Ribosomal L4; 1.
DR PROSITE; PS00539; RIBOSOMAL_L1E; 1.
DR Ribosomal protein.
SQ SEQUENCE 386 AA; 43135 MW; 23811BEEF015DD99 CRC64;

Query Match 72.7%; Score 32; DB 1; Length 386;
Best Local Similarity 66.7%; Pred. No. 39;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 STEVEGAGR 9
DB 360 APPVEGAGK 368

RESULT 5
APB_CHICK
ID APB_CHICK STANDARD; PRT; 433 AA.
AC P11882;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Apolipoprotein B (Fragment).
GN APOB.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88127960; PubMed=3436530;
RA Kirchgessner T.G., Heinemann C., Svenson K.L., Gordon D.A.,
RA Nicolsia M., Leberitz H.G., Lusis A.J., Williams D.L.;
RA "Regulation of chicken apolipoprotein B: cloning, tissue

distribution, and estrogen induction of mRNA.";
Gene 59:241-251(1987).
CC -!- FUNCTION: APOLIPOPROTEIN B IS A MAJOR PROTEIN CONSTITUENT OF
CHYLOMICRONS, VLDL AND LDL. IT FUNCTIONS AS A RECOGNITION
SIGNAL FOR THE CELLULAR BINDING AND INTERNALIZATION OF LDL
PARTICLES BY THE APOB/E RECEPTOR.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- INDUCTION: WITHIN 24 HOURS AFTER ESTRADIOL ADMINISTRATION,
APOB MRNA IS INCREASED FIVE- TO SEVEN-FOLD IN LIVER BUT IS
UNCHANGED IN INTESTINE AND KIDNEY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; M18421; AAA48595.1; -.
DR PIR; A29626; A29626.
KW Plasma; Lipid transport; VLDL; LDL; Chylomicron; Heparin-binding;
KW Cholesterol metabolism; Glycoprotein; Atherosclerosis.
FT NON TER 1
FT CARBOHYD 158 158 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 433 AA; 50847 MW; FD8808C9CFF48925 CRC64;

Query Match 72.7%; Score 32; DB 1; Length 433;
Best Local Similarity 75.0%; Pred. No. 44;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 TEVEGAGR 9
DB 253 TDVEGKGR 260

RESULT 6
SYFB_BRUME
ID SYFB_BRUME STANDARD; PRT; 804 AA.
AC Q8Y74;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Phenylalanine--tRNA synthetase beta chain (EC 6.1.1.20)
DE (Phenylalanine--tRNA ligase beta chain) (PHERS).
GN PHET OR BMEI2004.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29459;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=16M / ATCC 23456 / Biotype 1;
RX MEDLINE=20020109; PubMed=11756488;
RA DelVecchio V.G., Kapatral V., Redkar R.J., Patra G., Mijer C., Los T.,
RA Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Renik G.,
RA Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur G., Goldsman E.,
RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
RA Haselkorn R., Kyrpides N., Overbeek R.;
RA "The genome sequence of the facultative intracellular pathogen
Brucella melitensis.";
Proc Natl. Acad. Sci. U.S.A. 99:443-448(2002).
CC -!- CATALYTIC ACTIVITY: ATP + L-phenylalanine + tRNA(Phe) = AMP +
diphosphate + L-phenylalanyl-tRNA(Phe).
CC -!- COFACTOR: Binds 2 magnesium ions per tetramer (By similarity).
CC -!- SUBUNIT: Tetramer of two alpha and two beta chains (By
similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the phenylalanyl-tRNA synthetase beta chain
family. Subfamily 1.
CC -!- SIMILARITY: Contains 1 tRNA-binding domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 16:54:27 ; Search time 4.66667 Seconds
(without alignments)
90.694 Million cell updates/sec

Title: US-10-014-658-10

Perfect score: 44

Sequence: 1 STEVEGAGR 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	34	77.3	197	1	PCXA_BURCE
2	33	75.0	423	1	WR58 ARATH
3	32	72.7	324	1	E2B2_PYRHO
4	32	72.7	386	1	RL4_URECA
5	32	72.7	433	1	APB_CHICK
6	32	72.7	804	1	SYFB BRUSE
7	32	72.7	804	1	SYFB BRUSE
8	32	72.7	1167	1	CLJ4_BACTU
9	31	70.5	127	1	SECE_ECOLI
10	31	70.5	127	1	SECE_ECOLI
11	31	70.5	405	1	PR58_YEAST
12	31	70.5	429	1	HPAG_ECOLI
13	31	70.5	459	1	ZPR1_HUMAN
14	31	70.5	459	1	ZPR1_MOUSE
15	31	70.5	679	1	TKTA_CRAPL
16	31	70.5	1357	1	POB_PSEPU
17	31	70.5	1461	1	PRAX_HUMAN
18	30	68.2	112	1	GLNB_ECOLI
19	30	68.2	112	1	GLNB_KLEPN
20	30	68.2	112	1	GLNB_RHET
21	30	68.2	112	1	GLNB_RHOSH
22	30	68.2	112	1	GLNK_ECOLI
23	30	68.2	115	1	GLN2_METTH
24	30	68.2	121	1	Y554_XYLEA
25	30	68.2	229	1	HISA_PYRAE
26	30	68.2	250	1	THYA_CORGL
27	30	68.2	266	1	SUHB_RHIME
28	30	68.2	309	1	ARCL_MYCPN
29	30	68.2	312	1	ARC2_ENTFA
30	30	68.2	332	1	PYRB_CAUCR
31	30	68.2	428	1	R43F_MELGA
32	30	68.2	433	1	RHO_CHRVI
33	30	68.2	445	1	AMPA_MYCPN

Q99190 mus musculu
O53677 mycobacteri
P59040 chlorobium
P36721 human papil
O9hb58 homo sapien
P33148 xenopus lae
P55202 anguilla ja
P12021 sus scrofa
O70507 mus musculu
O60841 homo sapien
O63425 rattus norv
O55103 mus musculu

34 30 68.2 462 1 MCS1_MOUSE
35 30 68.2 494 1 COBQ_MYCTU
36 30 68.2 565 1 PYRG_CHLTE
37 30 68.2 605 1 VE1_HPV14
38 30 68.2 689 1 SP11_HUMAN
39 30 68.2 880 1 CADF_XENLA
40 30 68.2 1050 1 ANPE_ANGJA
41 30 68.2 1150 1 APMU_PIG
42 30 68.2 1186 1 HCN4_MOUSE
43 30 68.2 1220 1 IF2P_HUMAN
44 30 68.2 1383 1 PRAX_RAT
45 30 68.2 1391 1 PRAX_MOUSE

ALIGNMENTS

RESULT 1
PCXA_BURCE STANDARD; PRT; 197 AA.
AC P15109;1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-APR-1990 (Rel. 38, Last annotation update)
DE Protocatechuete 3,4-dioxygenase alpha chain (EC 1.13.11.3) (3,4-PCD).
GN PCAG.
OS Burkholderia cepacia (Pseudomonas cepacia).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
CC Burkholderiaceae; Burkholderia.
OX NCBI_TaxID=292;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90036674; PubMed=2808303;
RA Zylstra G.J., Olsen R.H., Ballou D.P.;
RT "Genetic organization and sequence of the Pseudomonas cepacia genes
for the alpha and beta subunits of protocatechuete 3,4-dioxygenase.";
RL J. Bacteriol. 171:5915-5921(1989).
CC -!- FUNCTION: PLAYS AN ESSENTIAL ROLE IN THE UTILIZATION OF NUMEROUS
AROMATIC AND HYDROAROMATIC COMPOUNDS VIA THE BETA-KETOADIPEATE
PATHWAY.
CC -!- CATALYTIC ACTIVITY: 3,4-dihydroxybenzoate + O(2) = 3-carboxy-
cis,cis-muconate.
CC -!- COFACTOR: FERRIC ION. EACH PROTEOMER APPEARS TO CONTAIN 1 ATOM OF
FERRIC IRON, FORMING THE ACTIVE SITE OF THE ENZYME.
CC -!- PATHWAY: SECOND STEP IN THE CONVERSION OF P-HYDROXYBENZOATE TO
SUCCINATE AND ACETYL-COA.
CC -!- SUBUNIT: THE ENZYME IS AN OLIGOMER OF 12 COPIES OF THE ALPHA AND
BETA CHAINS.
CC -!- SIMILARITY: BELONGS TO THE INTRADIOL RING-CLEAVAGE DIOXYGENASE
FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; M30791; AAA25925.1; -;
PIR; A33487; A33487.
HSSP; P00436; 3PCP.
InterPro: IPR000627; Dioxygenase.
Pfam; PF00775; Dioxygenase; 1.
PROSITE; PS00083; INTRADIOL DIOXYGENAS; 1.
KW Aromatic hydrocarbons catabolism; Oxidoreductase; Dioxygenase; Iron.
BINDING 130 130 PROTOCATHECHUATE (POTENTIAL).
SQ SEQUENCE 197 AA; 21394 MW; A00C1BCD420ECFC0 CRC64;

Query Match 77.3%; Score 34; DB 1; Length 197;
Best Local Similarity 75.0%; Pred.No. 7.5;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1";
RL Science 286:1571-1577(1999).
DR EMBL; AF001901; AAF09987.1; -.
DR TIGR; DR0407; -.
DR InterPro; IPR003544; CytC_biog_CcmB.
DR Pfam; PF03379; CcmB; 1.
DR PRINTS; PR01414; CCMBBIOGNIS.
KW Complete proteome.
SQ SEQUENCE 221 AA; 23518 MW; 84F40DAB73F1E0EC CRC64;

Query Match 72.7%; Score 32; DB 16; Length 221;
Best Local Similarity 75.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 STAVFFAG 8
|||:||||
DB 23 STLFVAG 30

RESULT 12

Q8DG36 PRELIMINARY; PRT; 357 AA.
AC Q8DG36;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE ABC-type transport system, permealase component.
GN VVI0011.
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=672;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CMCP6;
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RA Choy H.E.;
RT "Complete genome sequence of Vibrio vulnificus CMCP6";
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF016797; AAO08556.1; -.
KW Complete proteome.
SQ SEQUENCE 357 AA; 38162 MW; F8E1FC656569E570 CRC64;

Query Match 72.7%; Score 32; DB 16; Length 357;
Best Local Similarity 75.0%; Pred. No. 2e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 TAVFFAGR 9
|||:||||
DB 205 TAVFFAGR 212

RESULT 13

Q8KKB7 PRELIMINARY; PRT; 474 AA.
AC Q8KKB7;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE ORF.12 (Hypothetical protein).
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2205327; PubMed=12057956;
RA Raymond C.K., Sims E.H., Kas A., Spencer D.H., Kutyavin T.V.,
RA Ivey R.G., Zhou Y., Kaul R., Clendenning J.B., Olson M.V.;

Query Match 72.7%; Score 32; DB 16; Length 474;
Best Local Similarity 75.0%; Pred. No. 2e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 TAVFFAGR 9
|||:||||
DB 205 TAVFFAGR 212

RT "Genetic Variation at the O-Antigen Biosynthetic Locus in Pseudomonas
RT aeruginosa";
RL J. Bacteriol. 184:3614-3622(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=1-60;
RA Spencer D., Kas A., Smith E., Raymond C., Sims E., Hastings M.,
RA Burns J., Kaul R., Olson M.;
RT "Whole Genome Sequence Variation Among Multiple Isolates of Pseudomonas
RT aeruginosa";
RL J. Bacteriol. 0:0-0(2003).
DR EMBL; AF498400; RAM27546.1; -.
DR EMBL; AF540990; AAO17399.1; -.
KW Hypothetical protein.
SQ SEQUENCE 474 AA; 52911 MW; 130D9FFD1B089C17 CRC64;

Query Match 72.7%; Score 32; DB 2; Length 474;
Best Local Similarity 77.8%; Pred. No. 2.6e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 STAVFFAGR 9
|||:||||
DB 76 STAVFFAGR 84

RESULT 14

O67160 PRELIMINARY; PRT; 511 AA.
AC O67160;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Major facilitator family transporter.
GN EMR8 OR AQ.1062.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aujay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus";
RL Nature 392:353-358(1998).
DR EMBL; AE000721; AAC07127.1; -.
DR InterPro; IPR004638; Efflux_EmrB.
DR InterPro; IPR007114; MFS.
DR InterPro; IPR001411; TCR_TetB.
DR PRINTS; PR01036; TCRTEB.
DR TIGRFAMs; TIGR00711; efflux_EmrB; 1.
DR PROSITE; PSS0850; MFS; 1.
KW Complete proteome.
SQ SEQUENCE 511 AA; 58153 MW; 24E617934AECDB2E CRC64;

Query Match 72.7%; Score 32; DB 16; Length 511;
Best Local Similarity 66.7%; Pred. No. 2.8e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 STAVFFAGR 9
|||:||||
DB 91 TTSSFFAGR 99

RESULT 15

O94LJ1 PRELIMINARY; PRT; 566 AA.
AC O94LJ1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)

```
Query Match 79.5%; Score 35; DB 5; Length 403;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 AVFFAGR 9
DB 370 AVFFAGR 376

RESULT 4
O95978 PRELIMINARY; PRT; 157 AA.
AC O95978;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-2003 (TrEMBLrel. 23, Last annotation update)
DE Vh1 protein precursor (Fragment).
GN Vh1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Peripheral blood;
RA Jox A., Zander T., Kueppers R., Irsch J., Kanzler H., Kornacker M.,
RA Bohlen H., Diehl V., Wolf J.;
RT "Absence of immunoglobulin in Hodgkin-Reed Sternberg cells of a
RT patient with mixed cellularity Hodgkin's disease is associated with
RT somatic mutations within the untranslated regions of rearranged and
RT class switch recombinated Ig genes.";
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ005570; CAA06599.1; -
DR HSSP; P01772; 2FB4.
DR InterPro; IPR007110; IG_LIKE.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; IG; 1.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS00835; IG_LIKE; 1.
KW Signal.
FT SIGNAL.
FT NON_TER 157 157 POTENTIAL.
FT SEQUENCE 157 AA; 17304 MW; 86996EDDA84D88B5 CRC64;

Query Match 75.0%; Score 33; DB 4; Length 157;
Best Local Similarity 75.0%; Pred. No. 57;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 TAVFFAGR 9
DB 110 TAVFFAGR 117

RESULT 5
Q8PW33 PRELIMINARY; PRT; 264 AA.
AC Q8PW33;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein MM1768.
GN MM1768.
OS Methanosarcina mazei (Methanosarcina frisia).
OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OC NCBI_TaxID=2209;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Goel / Go1 / ATCC BAA-199 / DSM 3647 / OCM 88;
RX MEDLINE=22120827; PubMed=12125824;
RA Deppenmeier U., Johann A., Hartsch T., Merkl R., Schmitz R.A.,
RA Martinez-Arias R., Henne A., Wietzer A., Baeumer S., Jacob C.,
```

```
RA Brueggemann H., Lienard T., Christmann A., Boencke M., Steckel S.,
RA Bhattacharyya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,
RA Fritz H.-J., Gottschalk G.;
RT "The genome of Methanosarcina mazei: evidence for lateral gene
RT transfer between Bacteria and Archaea.";
RL J. Mol. Microbiol. Biotechnol. 4:453-461(2002).
DR EMBL; AE013412; AM31464.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 264 AA; 27419 MW; 412D1S9D4F9597FA CRC64;

Query Match 75.0%; Score 33; DB 17; Length 264;
Best Local Similarity 75.0%; Pred. No. 93;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 STAVFFAG 8
DB 230 STAVFFAG 237

RESULT 6
Q91MVO PRELIMINARY; PRT; 373 AA.
AC Q91MVO;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE LSDV057 putative virion core protein.
GN LSDV057 OR LD057.
OS Lumpy skin disease virus (LSDV).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Capripoxvirus.
OC NCBI_TaxID=59509;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Neethling 2490;
RX MEDLINE=21329495; PubMed=11435593;
RA Tulman E.R., Afonso C.L., Lu Z., Zaak L., Kutish G.F., Rock D.L.;
RT "Genome of lumpy skin disease virus.";
RL J. Virol. 75:7122-7130(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Neethling 2490;
RA Tulman E.R., Afonso C.L., Lu Z., Zaak L., Kutish G.F., Rock D.L.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Neethling Warmbaths LW;
RA Kara P.D., Afonso C.L., Wallace D.B., Kutish G.F., Stipinovich C.,
RA Lu Z., Vreede F.T., Taljaard L.C.F., Zaak A., Viljoen G.J., Rock D.L.;
RT "Molecular characterization of the South African vaccine strain and
RT the field isolate of lumpy skin disease virus.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF325528; AAK85018.1; -
DR EMBL; AF409137; AAN02625.1; -
SQ SEQUENCE 373 AA; 42057 MW; F1AF112A183F43E4 CRC64;

Query Match 75.0%; Score 33; DB 12; Length 373;
Best Local Similarity 75.0%; Pred. No. 13e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 STAVFFAG 8
DB 190 STAVFFAG 197

RESULT 7
Q8JUT7 PRELIMINARY; PRT; 373 AA.
AC Q8JUT7;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Putative virion core protein.
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 18:03:30 ; Search time 20.3333 Seconds
(without alignments)
114.220 Million cell updates/sec

Title: US-10-014-658-15

Perfect score: 44

Sequence: 1 STAFFPAGR 9

Scoring table: BLCUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 259052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 23.1*

1: sp_archaea.*

2: sp_bacteria.*

3: sp_fungi.*

4: sp_human.*

5: sp_invertebrate.*

6: sp_mammal.*

7: sp_mhc.*

8: sp_organelle.*

9: sp_phage.*

10: sp_plant.*

11: sp_rodent.*

12: sp_virus.*

13: sp_vertebrate.*

14: sp_unclassified.*

15: sp_virus.*

16: sp_bacteriap.*

17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36	81.8	307	17 Q9HSY4	Q9HSY4 halobacteri
2	35	79.5	220	5 Q8T322	Q8T322 drosophila
3	35	79.5	403	5 Q9V989	Q9V989 drosophila
4	33	75.0	157	4 Q95978	Q95978 homo sapien
5	33	75.0	264	17 Q8PW33	Q8PW33 methanoea
6	33	75.0	373	12 Q91MVO	Q91MVO lumpy skin
7	33	75.0	373	12 Q8JTB7	Q8JTB7 lumpy skin
8	33	75.0	433	2 Q9PCQ6	Q9PCQ6 pseudomonas
9	32	72.7	121	10 Q9AMQ8	Q9AMQ8 oryza sativ
10	32	72.7	199	16 Q8EZ10	Q8EZ10 leptospira
11	32	72.7	221	16 Q9RXA8	Q9RXA8 deinococcus
12	32	72.7	357	16 Q8DG36	Q8DG36 vibrio vuln
13	32	72.7	474	2 Q8XNB7	Q8XNB7 pseudomonas
14	32	72.7	511	16 Q67160	Q67160 aquifex aeo
15	32	72.7	566	10 Q94LJ1	Q94LJ1 oryza sativ
16	32	72.7	642	16 Q8U9J0	Q8U9J0 agrobacteri

17	32	72.7	1013	2 Q85920	Q85920 sphingomona
18	32	72.7	1714	10 Q9FWM3	Q9FWM3 arabidopsis
19	31	70.5	55	4 Q9UC78	Q9UC78 homo sapien
20	31	70.5	57	4 Q8J001	Q8J001 homo sapien
21	31	70.5	58	4 Q9UBW9	Q9UBW9 homo sapien
22	31	70.5	58	4 Q8IZZ9	Q8IZZ9 homo sapien
23	31	70.5	59	4 Q8J000	Q8J000 homo sapien
24	31	70.5	90	10 Q8RZ00	Q8RZ00 oryza sativ
25	31	70.5	110	17 Q8Z252	Q8Z252 pyrobaculum
26	31	70.5	115	2 Q8RP93	Q8RP93 streptococc
27	31	70.5	115	16 Q87PU4	Q87PU4 streptococc
28	31	70.5	115	16 Q8DP51	Q8DP51 streptococc
29	31	70.5	131	5 Q9W0B7	Q9W0B7 drosophila
30	31	70.5	133	4 Q15908	Q15908 homo sapien
31	31	70.5	167	10 Q94ES1	Q94ES1 zea mays (m
32	31	70.5	193	2 Q07682	Q07682 escherichia
33	31	70.5	200	12 Q55482	Q55482 porcine rep
34	31	70.5	200	12 Q41187	Q41187 porcine rep
35	31	70.5	200	12 Q9J7K2	Q9J7K2 porcine rep
36	31	70.5	200	12 Q990Q8	Q990Q8 porcine rep
37	31	70.5	200	12 Q9ICCS	Q9ICCS porcine rep
38	31	70.5	200	12 Q55486	Q55486 porcine rep
39	31	70.5	200	12 Q55490	Q55490 porcine rep
40	31	70.5	200	12 Q41135	Q41135 porcine rep
41	31	70.5	200	12 Q9J7K4	Q9J7K4 porcine rep
42	31	70.5	200	12 Q9J7J5	Q9J7J5 porcine rep
43	31	70.5	233	10 Q48889	Q48889 zea mays (m
44	31	70.5	234	16 Q8UGF6	Q8UGF6 agrobacteri
45	31	70.5	239	8 Q9TL13	Q9TL13 nephroselm

ALIGNMENTS

RESULT 1

Q9HSY4 PRELIMINARY; PRT; 307 AA.
ID Q9HSY4
AC Q9HSY4
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE Vng0026C.
GN VNG0026C.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA NG W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Shrogha J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welter R., Goo Y.A.,
RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Izenbarger T.A., Peck R.P., Pohlschroder M., Spudis J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.,
RT "Genome sequence of Halobacterium species NRC-1";
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR EMBL; AB004971; AAC18666.1; -.
KW Complete proteome.
SQ SEQUENCE 307 AA; 34492 MW; AB92A26FF80AF39 CRC64;

Query Match 81.8%; Score 36; DB 17; Length 307;

Best Local Similarity 77.8%; Pred. No. 26;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 STAFFPAGR 9

Db 92 STAFFPAGR 100

```

CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE002286; AAF39070.1; -.
DR PIR; A81731; A81731.
DR TIGR; TC0197; -.
DR InterPro; IPR005546; Autotransporter.
DR InterPro; IPR003368; Chlamydia_PMP.
DR Pfam; PF03797; Autotransporter; 1.
DR Pfam; PF02415; DUF145; 1.
DR TIGRFAMs; TIGR01376; POMP_repeat; 15.
DR Outer membrane; Signal; Multigene family; Complete proteome.
KW SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 1520 PROBABLE OUTER MEMBRANE PROTEIN PMED.
SQ SEQUENCE 1520 AA; 162102 MW; 6881405AC79585F CRC64;

Query Match 70.5%; Score 31; DB 1; Length 1520;
Best Local Similarity 55.6%; Pred. No. 1.5e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 STAVFPAGR 9
DB 920 SDAYFAGK 928

```

Search completed: December 11, 2003, 18:21:20
 Job time : 6.66667 secs

RT subtypes.", Acad. Sci. U.S.A. 92:9373-9377(1995).
RL FUNCTION: HIGH-AFFINITY H+/SULFATE COTRANSPORTER THAT MEDIATES THE
CC UPTAKE OF SULFATE BY PLANT ROOTS FROM LOW CONCENTRATIONS OF
CC SULFATE IN THE SOIL SOLUTION.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE SLC26A FAMILY OF TRANSPORTERS.
CC -1- SIMILARITY: Contains 1 STAS domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X82256; CAAS7711.1; -
CC PIR: S51764; S51764.
CC InterPro: IPR002645; STAS.
CC InterPro: IPR001902; Sulph_transpt.
CC Pfam: PF01740; STAS; 1.
CC TIGRPFAM: TIGR00815; sulph; 1.
CC PROSITE: PS01130; SLC26A; 1.
CC PROSITE: PS0801; STAS; 1.
CC Transmembrane; Transport; Multigene family.
CC FT TRANSMEM 103 123 POTENTIAL.
CC FT TRANSMEM 128 148 POTENTIAL.
CC FT TRANSMEM 153 173 POTENTIAL.
CC FT TRANSMEM 182 202 POTENTIAL.
CC FT TRANSMEM 205 225 POTENTIAL.
CC FT TRANSMEM 264 284 POTENTIAL.
CC FT TRANSMEM 291 311 POTENTIAL.
CC FT TRANSMEM 345 366 POTENTIAL.
CC FT TRANSMEM 383 403 POTENTIAL.
CC FT TRANSMEM 420 440 POTENTIAL.
CC FT TRANSMEM 447 467 POTENTIAL.
CC FT TRANSMEM 481 501 POTENTIAL.
CC FT DOMAIN 532 655 STAS.
CC SQ SEQUENCE 662 AA; 72728 MW; 7544D1E73FB76284 CRC64;
Query Match 70.5%; Score 31; DB 1; Length 662;
Best Local Similarity 85.7%; Pred. No. 68;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 TAVFFAG 8
DB 187 TATFFAG 193
RESULT 12
SUTI_STYHA STANDARD; PRT; 667 AA.
AC P53391;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE High affinity sulphate transporter 1.
GN STL.
OS Stylosanthes hamata (Caribbean stylo).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids I; Fabales; Fabaceae; Papilionoideae; Aeschynomeneae;
OC Stylosanthes
CC NCBI_TaxID=37660;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Verano; TISSUE=Root;
RX MEDLINE=96016171; PubMed=7568135;
RA Smith F.W., Ealing P.M., Hawkesford M.J., Clarkson D.T.;
RT "Plant members of a family of sulfate transporters reveal functional
subtypes.";

Proc. Natl. Acad. Sci. U.S.A. 92:9373-9377(1995).
RL FUNCTION: HIGH-AFFINITY H+/SULFATE COTRANSPORTER THAT MEDIATES THE
CC UPTAKE OF SULFATE BY PLANT ROOTS FROM LOW CONCENTRATIONS OF
CC SULFATE IN THE SOIL SOLUTION.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE SLC26A FAMILY OF TRANSPORTERS.
CC -1- SIMILARITY: Contains 1 STAS domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X82255; CAAS7710.1; -
CC PIR: S51763; S51763.
CC InterPro: IPR002645; STAS.
CC InterPro: IPR001902; Sulph_transpt.
CC Pfam: PF01740; STAS; 1.
CC TIGRPFAM: TIGR00815; sulph; 1.
CC PROSITE: PS01130; SLC26A; 1.
CC PROSITE: PS0801; STAS; 1.
CC Transmembrane; Transport; Multigene family.
CC FT TRANSMEM 106 126 POTENTIAL.
CC FT TRANSMEM 131 151 POTENTIAL.
CC FT TRANSMEM 156 176 POTENTIAL.
CC FT TRANSMEM 185 205 POTENTIAL.
CC FT TRANSMEM 208 228 POTENTIAL.
CC FT TRANSMEM 269 289 POTENTIAL.
CC FT TRANSMEM 296 316 POTENTIAL.
CC FT TRANSMEM 350 370 POTENTIAL.
CC FT TRANSMEM 425 445 POTENTIAL.
CC FT TRANSMEM 452 472 POTENTIAL.
CC FT TRANSMEM 486 506 POTENTIAL.
CC FT DOMAIN 537 660 STAS.
CC SQ SEQUENCE 667 AA; 73172 MW; 1A7AD47FDEB4DA7E CRC64;
Query Match 70.5%; Score 31; DB 1; Length 667;
Best Local Similarity 85.7%; Pred. No. 68;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 TAVFFAG 8
DB 190 TATFFAG 196
RESULT 13
SYL_THEAC STANDARD; PRT; 910 AA.
AC Q9HK31;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LeuRS).
GN LEUS OR TA0777.
OS Thermoplasma acidophilum.
OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
OC Thermoplasmataceae; Thermoplasma.
CC NCBI_TaxID=2303;
OX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 1728;
RX MEDLINE=20479972; PubMed=11029001;
RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
RA Newes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma
acidophilum.";
RT Nature 407:508-513(2000).
CC -1- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC

- RA Olds R.J., Lane D.A., Chowdhury V., de Stefano V., Leone G.,
RA Thein S.L.;
RT "Complete nucleotide sequence of the antithrombin gene: evidence for
RT homologous recombination causing thrombophilia.";
RL Biochemistry 32:4216-4224(1993).
[4]
RN SEQUENCE FROM N.A.
RP TISSUE=Fetal liver;
RA Zhang C., Yu Y., Zhang S., Wei H., Bi J., Zhou G., Dong C., Zai Y.,
RA Xu W., Gao F., Liu M., He F.;
RT "Functional prediction of the coding sequences of 75 new genes deduced
RT by analysis of cDNA clones from human fetal liver.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
[5]
RN SEQUENCE FROM N.A., AND VARIANTS GLU-30 AND ALA-147.
RP Rieder M.J., Carrington D.P., Chung M.-W., Lee K.L., Poel C.L., Yi Q.,
RA Nickerson D.A.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
[6]
RN SEQUENCE OF 42-464 FROM N.A.
RX MEDLINE=83238456; PubMed=6305982;
RA Prochownik E.V., Markham A.P., Orkin S.H.;
RT "Isolation of a cDNA clone for human antithrombin III.";
RL J. Biol. Chem. 258:8389-8394(1983).
[7]
RN SEQUENCE OF 33-464, CARBOHYDRATE-LINKAGE SITES, AND DISULFIDE BONDS.
RA Petersen T.E., Dudek-Wojciechowska G., Sottrup-Jensen L.,
RA Magnusson S.;
RT "Primary structure of antithrombin-III (heparin cofactor). Partial
RT homology between alpha-1-antitrypsin and antithrombin-III.";
RL (In) Collen D., Wiman B., Verstraete M. (eds.);
RL The physiological inhibitors of blood coagulation and fibrinolysis,
RL pp.43-54, Elsevier, Amsterdam (1979).
[8]
RN ACTIVE SITE.
RP MEDLINE=81212814; PubMed=7238875;
RX Bjoerk I., Danielsson A., Fenton J.W. II, Joernvall H.;
RA "The site in human antithrombin for functional proteolytic cleavage
RT by human thrombin.";
RL FEBS Lett. 126:257-260(1981).
[9]
RN HEPARIN-BINDING SITE.
RP MEDLINE=84111578; PubMed=6693405;
RX Blackburn M.N., Smith R.L., Carlson J., Sibley C.C.;
RA "The heparin-binding site of antithrombin III. Identification of a
RT critical tryptophan in the amino acid sequence.";
RL J. Biol. Chem. 259:939-941(1984).
[10]
RN MUTAGENESIS OF ALA-414.
RX MEDLINE=91192143; PubMed=2013320;
RA Austin R.C., Rachubinski R.A., Blachjman M.A.;
RT "Site-directed mutagenesis of alanine-382 of human antithrombin III.";
RL FEBS Lett. 280:254-258(1991).
[11]
RN X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).
RP MEDLINE=94373498; PubMed=8087553;
RX Carrell R.W., Stein P.E., Fermi G., Wardell M.R.;
RA "Biological implications of a 3 A structure of dimeric antithrombin.";
RL Structure 2:257-270(1994).
[12]
RN X-RAY CRYSTALLOGRAPHY (3.6 ANGSTROMS).
RX MEDLINE=95384753; PubMed=7656006;
RA Schreuder H.A., de Boer B., Dijkema R., Mulders J., Theunissen H.J.M.,
RA Grootehuys P.D.J., Hol W.G.J.;
RT "The intact and cleaved human antithrombin III complex as a model for
RT serpin-proteinase interactions.";
RL Nat. Struct. Biol. 1:48-54(1994).
[13]
RN X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
RX MEDLINE=97202038; PubMed=9067613;
RA Skinner R., Abrahams J.-P., Whistock J.C., Lesk A.M., Carrel R.W.,
RA Wardell M.R.;
RT "The 2.6 A structure of antithrombin indicates a conformational
- change at the heparin binding site.";
RL J. Mol. Biol. 266:601-609(1997).
[14]
RN X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS).
RX MEDLINE=98437386; PubMed=9761689;
RA Skinner R., Chang W.S.W., Jin L., Pei X.Y., Huntington J.A.,
RA Abrahams J.-P., Carrell R.W., Lomas D.A.;
RT "Implications for function and therapy of a 2.9 A structure of
RT binary-complexed antithrombin.";
RL J. Mol. Biol. 283:9-14(1998).
[15]
RN REVIEW.
RP MEDLINE=91129302; PubMed=2126464;
RX Mourey L., Samama J.-P., Delarue M., Choay J., Lormeau J.C.,
RA Petitou M., Moras D.;
RT "Antithrombin III: structural and functional aspects.";
RL Biochimie 72:599-608(1990).
[16]
RN REVIEW ON VARIANTS
RP MEDLINE=94054329; PubMed=8236149;
RX Lane D.A., Olds R.J., Boiscialar M., Chowdhury V., Thein S.L.,
RA Cooper D.N., Blachman M., Perry D., Emmerich J., Alach M.;
RT "Antithrombin III mutation database: first update. For the Thrombin
RT and its Inhibitors Subcommittee of the Scientific and Standardization
RT Committee of the International Society on Thrombosis and
RT Haemostasias.";
RL Thromb. Haemost. 70:361-369(1993).
[17]
RN REVIEW ON VARIANTS.
RP MEDLINE=95269065; PubMed=7749926;
RX Stein P.E., Carrell R.W.;
RT "What do dysfunctional serpins tell us about molecular mobility and
RT disease?";
RL Nat. Struct. Biol. 2:96-113(1995).
[18]
RN REVIEW ON VARIANTS.
RP MEDLINE=96263733; PubMed=8664906;
RX Perry D.J., Carrell R.W.;
RT "Molecular genetics of human antithrombin deficiency.";
RL Hum. Mutat. 7:7-22(1996).
[19]
RN VARIANT UTAH.
RP MEDLINE=89050967; PubMed=3191114;
RX Bock S.C., Marrinan J.A., Radziejewska E.;
RA "Antithrombin III Utah: proline-407 to leucine mutation in a highly
RT conserved region near the inhibitor reactive site.";
RL Biochemistry 27:6171-6178(1988).
[20]
RN VARIANT TOYAMA.
RP MEDLINE=84119472; PubMed=6592486;
RX Koide T., Ocani S., Takanashi K., Ono T., Sakuragawa N.;
RA "Antithrombin III Toyama: replacement of arginine-47 by cysteine in
RT hereditary abnormal antithrombin III that lacks heparin-binding
RT ability.";
RL Proc. Natl. Acad. Sci. U.S.A. 81:289-293(1984).
[21]
RN VARIANT BASEL.
RP MEDLINE=86111754; PubMed=3080419;
RX Chang J.Y., Tran T.H.;
RT "Antithrombin III Basel. Identification of a Pro-Leu substitution in
RT a hereditary abnormal antithrombin with impaired heparin cofactor
RT activity.";
RL J. Biol. Chem. 261:1174-1176(1986).
[22]
RN VARIANT DENVER.
RX MEDLINE=87109210; PubMed=3805013;
RA Stephens A.W., Thalley B.S., Hirs C.H.W.;
RT "Antithrombin-III Denver, a reactive site variant.";
RL J. Biol. Chem. 262:1044-1048(1987).
[23]
RN VARIANT HAMILTON.
RP MEDLINE=89027076; PubMed=3179438;
RX Devrak-Kizuk R., Chui D.H.K., Prochownik E.V., Carter C.J.,


```

RX MEDLINE=97124195; PubMed=8969508;
RA Mizuno M., Masuda S., Takemaru K.-I., Hosono S., Sato T., Takeuchi M.,
RA Kobayashi Y.;
RT "Systematic sequencing of the 283 kb 210 degrees-232 degrees region of
RT the Bacillus subtilis genome containing the skin element and many
RT sporulation genes.";
RL Microbiology 142:3103-3111(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunat F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borviss R., Bourcier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Comerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kashara Y., Kassar-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portelle B., Porcellik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rappoport G., Rey M., Reynolds S.,
RA Rieder M., Rivolta C., Rocha E., Roche J., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Scrocin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier P., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzneger T.,
RA Winers P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256(1997).
RN [4]
RP IDENTIFICATION
RX MEDLINE=96084975; PubMed=7489895;
RA Medigue C., Moszer I., Viari A., Danchin A.;
RT "Analysis of a Bacillus subtilis genome fragment using a co-operative
RT computer system prototype.";
RL Gene 165:GC37-GC51(1995).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; D32216; BAA06931.1; -
DR EMBL; D84432; BAA12393.1; -
DR EMBL; D99117; CAB14562.1; -
DR PIR; A69946; A69946.
DR Subtilist; BG11269; yqar.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 154 AA; 18112 MW; CDBAF5E82B607B7B CRC64;

Query Match 70.5%; Score 31; DB 1; Length 154;
Best Local Similarity 85.7%; Pred. No. 17;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TAVFFAG 8
Db 15 TAVFTG 21

RX MEDLINE=97124195; PubMed=8969508;
RA Mizuno M., Masuda S., Takemaru K.-I., Hosono S., Sato T., Takeuchi M.,
RA Kobayashi Y.;
RT "Systematic sequencing of the 283 kb 210 degrees-232 degrees region of
RT the Bacillus subtilis genome containing the skin element and many
RT sporulation genes.";
RL Microbiology 142:3103-3111(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunat F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borviss R., Bourcier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Comerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kashara Y., Kassar-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portelle B., Porcellik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rappoport G., Rey M., Reynolds S.,
RA Rieder M., Rivolta C., Rocha E., Roche J., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Scrocin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier P., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzneger T.,
RA Winers P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256(1997).
RN [4]
RP IDENTIFICATION
RX MEDLINE=96084975; PubMed=7489895;
RA Medigue C., Moszer I., Viari A., Danchin A.;
RT "Analysis of a Bacillus subtilis genome fragment using a co-operative
RT computer system prototype.";
RL Gene 165:GC37-GC51(1995).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; D32216; BAA06931.1; -
DR EMBL; D84432; BAA12393.1; -
DR EMBL; D99117; CAB14562.1; -
DR PIR; A69946; A69946.
DR Subtilist; BG11269; yqar.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 154 AA; 18112 MW; CDBAF5E82B607B7B CRC64;

Query Match 70.5%; Score 31; DB 1; Length 154;
Best Local Similarity 85.7%; Pred. No. 17;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TAVFFAG 8
Db 15 TAVFTG 21

```

```

RESULT 5
YRBE_HAEIN STANDARD; PRT; 261 AA.
AC P45030;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein H11086.
GN H11086.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellales; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RC MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Karlayage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RA "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -1- SIMILARITY: STRONG, TO E.COLI YRBE.
CC -1- SIMILARITY: TO P.PURPUREA HYPOTHETICAL 28.3 kDa PROTEIN IN YCF26-
CC CHLI INTERGENIC REGION (AC P51393).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; U52788; AAC22742.1; -
DR PIR; D64166; D64166.
DR TIGR; H11086; -
DR InterPro; IPR003453; DUF140.
DR Pfam; PF02405; DUF140; 1.
DR TIGRfam; TIGR00056; TIGR00056; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 13 33 POTENTIAL.
FT TRANSMEM 50 70 POTENTIAL.
FT TRANSMEM 90 110 POTENTIAL.
FT TRANSMEM 148 168 POTENTIAL.
FT TRANSMEM 199 219 POTENTIAL.
FT TRANSMEM 239 259 POTENTIAL.
SQ SEQUENCE 261 AA; 28015 MW; 4BC3695F247A6BF6 CRC64;

Query Match 70.5%; Score 31; DB 1; Length 261;
Best Local Similarity 75.0%; Pred. No. 28;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 TAVFFAG 9
Db 104 TALLFAG 111

RESULT 6
HTPX_RALSO STANDARD; PRT; 286 AA.
ID_HTPX_RALSO
AC Q8Y3A6;
DT 28-FEB-2003 (Rel. 41, Created)

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 16:54:27 ; Search time 4.66667 Seconds
(without alignments)
90.694 Million cell updates/sec

Title: US-10-014-658-15
Perfect score: 44
Sequence: 1 STAFFFAGR 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	35	79.5	216	1 GSPJ_ERWCA	P31589 erwinia car
2	35	79.5	282	1 Y125_THEME	Q9wx99 thermotoga
3	33	75.0	198	1 GSPJ_KLEPN	P15749 klebsiella
4	31	70.5	154	1 YQAR_BACSU	P45914 bacillus su
5	31	70.5	261	1 YRBE_HAEIN	P45030 haemophilus
6	31	70.5	286	1 HTEK_RALSO	Q892a6 raistonia s
7	31	70.5	423	1 PUR2_RHIME	Q92r10 rhizobium m
8	31	70.5	464	1 ANT3_HUMAN	P01008 homo sapien
9	31	70.5	476	1 VL2_HPV32	P36757 human papil
10	31	70.5	521	1 VL2_HPV49	P36762 human papil
11	31	70.5	562	1 SUT2_STYHA	P53392 stylosanthe
12	31	70.5	667	1 SUT1_STYHA	P53391 stylosanthe
13	31	70.5	910	1 SYL_THEAC	Q9hk31 thermoplas
14	31	70.5	941	1 SYL_HAEIN	P43824 haemophilus
15	31	70.5	1520	1 PMPD_CHLMU	Q9plb0 chlamydia m
16	31	70.5	1531	1 PMPD_CHLTR	Q84818 chlamydia t
17	30	68.2	119	1 N4BM_HUMAN	Q95298 homo sapien
18	30	68.2	131	1 YJGH_ECOLI	P39332 escherichia
19	30	68.2	163	1 DUT_AEPEE	Q9vg32 aeropyrum p
20	30	68.2	189	1 EFP_PSEPK	Q88180 pseudomonas
21	30	68.2	285	1 PSTB_MYCIT	Q49588 mycobacteri
22	30	68.2	527	1 VL2_HPV38	Q80912 human papil
23	30	68.2	533	1 VL2_HPV15	P36750 human papil
24	30	68.2	550	1 PUTX_EMENI	P18696 emericepia
25	30	68.2	660	1 MM02_HUMAN	P08253 homo sapien
26	30	68.2	662	1 MM02_MOUSE	P33434 mus musculus
27	30	68.2	662	1 MM02_RABIT	P50757 oryctolagus
28	30	68.2	662	1 MM02_RAT	P33436 rattus norv
29	30	68.2	663	1 MM02_CHICK	Q90611 gallus gall
30	30	68.2	1762	1 POLN_FCVG6	P27407 feline cali
31	29	65.9	169	1 Y358_BUCAI	P57439 buchnera ap
32	29	65.9	250	1 BTG4_MOUSE	O70552 mus musculu
33	29	65.9	257	1 ZUPT_SALTY	Q8xgr4 salmonella

ALIGNMENTS

RESULT 1

GSPJ_ERWCA	STANDARD;	PRT;	216 AA.
ID	GSPJ_ERWCA		
AC	P31589;		
DT	01-JUL-1993 (Rel. 26, Created)		
DT	01-JUL-1993 (Rel. 26, Last sequence update)		
DT	01-JUL-1993 (Rel. 26, Last annotation update)		
DE	General secretion pathway protein J precursor (pectic enzymes secretion protein outJ).		
GN	OUTJ.		
OS	Erwinia carotovora.		
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;		
OC	Enterobacteriaceae; Pectobacterium.		
OX	NCBI_TaxID=554;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=SCRI193;		
RX	MEDLINE=93316842; PubMed=8326859;		
RA	Reeves P.J., Whitcombe D., Wharam S., Gibson M., Allison G., Bunce N., Barallion R., Douglas P., Mulholland V., Stevens S., Walker S., Salmund G.P.C.;		
RA	"Molecular cloning and characterization of 13 out genes from Erwinia carotovora subspecies carotovora: genes encoding members of a general secretion pathway (GSP) widespread in Gram-negative bacteria.";		
RL	Mol. Microbiol. 8:443-456(1993).		
CC	!- FUNCTION: INVOLVED IN A GENERAL SECRETION PATHWAY (GSP) FOR THE EXPORT OF PROTEINS. REQUIRED FOR THE TRANSLOCATION OF THE MULTIPLE PECTIC ENZYMES.		
CC	!- SIMILARITY: BELONGS TO THE PULJ/OUTJ/XPSJ/EXEJ/XCPW FAMILY.		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).		
CC	EMBL; X70049; CAA49651.1; -		
DR	PIR; S32864; S32864.		
DR	InterPro; IPR003413; GSPJ1_IJ.		
DR	InterPro; IPR001120; Prok_N_methyltn.		
DR	Pfam; PF02501; GSPJ1_IJ_1.		
DR	PROSITE; PS00409; PROKAR_NTER_METHYL; 1.		
DR	Transprot; Methylation.		
KW	PROPEP 1 25		
FT	CHAIN 26 216		
FT	MOD RES 26 216		
FT	SEQUENCE 216 AA; 24850 MW; 96223BCA29471070 CRC64;		

Query Match 79.5%; Score 35; DB 1; Length 216;
Best Local Similarity 66.7%; Pred. No. 3.5;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
QY 1 STAFFFAGR 9
|||::|||

Q08632 picea abies
Q8k9i7 streptococc
Q8p3d2 streptococc
Q9alb6 streptococc
Q69861 streptomyce
Q44775 borrelia bu
P27234 human papil
P22739 xenopus lae
P54925 bartonella
P40831 m glucosami
Q92JJ9 rickettsia
Q9zeaz rickettsia

```
Query Match      72.7%; Score 32; DB 1; Length 511;
Best Local Similarity 66.7%; Pred. No. 85;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 STAVFFAGR 9
   ||:|||||
Db 91 TTSSFFAGR 99

RESULT 8
AF3016
NA+/H+ antiporter Atu3738 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C/Species: Agrobacterium tumefaciens
C/Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C/Accession: AF3016
R/Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.;
Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A/Reference number: AB2577; MUID:21608550; PMID:11743193
A/Accession: AF3016
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-621 <KUR>
A/Cross-references: GB:AE008689; PIDN:ARL44548.1; PID:gl7742162; GSPDB:GN00187
A/Experimental source: strain C58 (Dupont)
C/Genetics:
A/Gene: Atu3738
A/Map position: linear chromosome

Query Match      72.7%; Score 32; DB 2; Length 621;
Best Local Similarity 85.7%; Pred. No. 1e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 TAVFFAG 8
   ||:|||||
Db 100 TSVFFAG 106

RESULT 9
B98268
probable sodium/hydrogen antiporter PA5021 [imported] - Agrobacterium tumefaciens (strain
C/Species: Agrobacterium tumefaciens
C/Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C/Accession: B98268
R/Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A/Reference number: A97359; MUID:21608551; PMID:11743194
A/Accession: B98268
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-642 <KUR>
A/Cross-references: GB:AE007870; PIDN:AAK89668.1; PID:gl5159570; GSPDB:GN00170
C/Genetics:
A/Gene: AGR_L_2194
A/Map position: linear chromosome

Query Match      72.7%; Score 32; DB 2; Length 642;
Best Local Similarity 85.7%; Pred. No. 1.1e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 TAVFFAG 8
   ||:|||||
Db 121 TSVFFAG 127

RESULT 10
T31211
```

```
trwC protein homolog - Sphingomonas aromaticivorans plasmid pNL1
C/Species: Sphingomonas aromaticivorans
C/Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 11-Jan-2000
C/Accession: T31211
R/Romine, M.F.; Stillwell, L.C.; Wong, K.K.; Thurston, S.J.; Sisk, E.C.; Sensen, C.W.;
submitted to the EMBL Data Library, July 1998
A/Description: Complete sequence of a 184 kb catabolic plasmid from Sphingomonas aromati
A/Reference number: Z20992
A/Accession: T31211
A/Status: preliminary; translated from GB/EMBL/DBDJB
A/Molecule type: DNA
A/Residues: 1-1013 <ROM>
A/Cross-references: EMBL:AF079317; MID:g3378261; PID:g3378352; PIDN:AAD03935.1
C/Genetics:
A/Genome: plasmid pNL1
A/Note: orf704

Query Match      72.7%; Score 32; DB 2; Length 1013;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 STAVFFAGR 9
   ||:|||||
Db 705 ATAFTAGR 713

RESULT 11
E95174
bacteriocin transport accessory protein SP1499 [imported] - Streptococcus pneumoniae (str
C/Species: Streptococcus pneumoniae
C/Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C/Accession: E95174
R/Tattelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Hei
on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel,
son, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A/Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A/Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae
A/Reference number: A95000; MUID:21357209; PMID:11463916
A/Accession: E95174
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-115 <KUR>
A/Cross-references: GB:AE005672; PIDN:AAK75590.1; PID:gl4972988; GSPDB:GN00164; TIGR:SP4
A/Experimental source: strain TIGR4
C/Genetics:
A/Gene: SP1499

Query Match      70.5%; Score 31; DB 2; Length 115;
Best Local Similarity 75.0%; Pred. No. 35;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TAVFFAGR 9
   ||:|||||
Db 28 TATFFIGR 35

RESULT 12
G98040
bacteriocin transport accessory protein [imported] - Streptococcus pneumoniae (strain R6)
C/Species: Streptococcus pneumoniae
C/Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C/Accession: G98040
R/Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; F
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, F.; McAhren, S.; N
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A/Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A/Title: Genome of the bacterium Streptococcus pneumoniae Strain R6.
A/Reference number: A97872; MUID:21429245; PMID:11544234
A/Accession: G98040
A/Status: preliminary
A/Molecule type: DNA
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 18:19:54 ; Search time 8.3333 Seconds
(without alignments)
103.862 Million cell updates/sec

Title: US-10-014-659-15

Perfect score: 44

Sequence: 1 STAVFPFAGR 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Pir1.*

2: Pir2.*

3: Pir3.*

4: Pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	36	81.8	307	F84162	hypothetical prote
2	35	79.5	216	S32864	outJ protein - Erw
3	35	79.5	282	E72415	zinc ABC transport
4	33	75.0	126	S58121	Ig heavy chain V r
5	33	75.0	198	S11920	pulJ protein - Kle
6	32	72.7	221	H75521	cytochrome c-type
7	32	72.7	511	E70391	major facilitator
8	32	72.7	621	AF3016	Na+/H+ antiporter
9	32	72.7	642	B98268	probable sodium/hy
10	32	72.7	1013	T31211	trwC protein homol
11	31	70.5	115	E95174	bacteriocin transpo
12	31	70.5	115	G98040	bacteriocin transpo
13	31	70.5	154	A69946	hypothetical prote
14	31	70.5	233	T01205	sulfate transport
15	31	70.5	234	A12705	branched-chain ami
16	31	70.5	234	B37492	azlC family protei
17	31	70.5	258	A82068	conserved hypothet
18	31	70.5	260	AG0434	probable membrane
19	31	70.5	261	D64166	hypothetical prote
20	31	70.5	294	T10445	peroxidase (EC 1.1
21	31	70.5	307	T18990	hypothetical prote
22	31	70.5	378	A84161	hypothetical prote
23	31	70.5	404	T16272	hypothetical prote
24	31	70.5	464	1 XHHU3	antithrombin III p
25	31	70.5	476	1	L2 protein - human
26	31	70.5	521	2 T48587	peptide transporte
27	31	70.5	624	2 T48902	sulfate transporte
28	31	70.5	646	2 T01079	sulfate transport
29	31	70.5	646	2	

30 31 70.5 649 2 T51839 sulfate transport
31 31 70.5 649 2 T00946 probable sulfate t
32 31 70.5 658 2 T49001 sulfate transporte
33 31 70.5 660 2 T49069 sulfate transport
34 31 70.5 662 2 S51764 sulfate transport
35 31 70.5 667 2 S51763 sulfate transport
36 31 70.5 671 2 D38490 matutase-related p
37 31 70.5 671 2 D38490 matutase-related p
38 31 70.5 703 2 T51161 hypothetrical prote
39 31 70.5 719 2 D96809 protein F28K19.22
40 31 70.5 757 2 T02884 D-(-)-3-hydroxybut
41 31 70.5 941 2 S78633 isoleucine-tRNA li
42 31 70.5 1520 2 A81731 polymorphic membra
43 31 70.5 1531 2 H71468 probable outer mem
44 30 68.2 119 2 T08728 NADH2 dehydrogenas
45 30 68.2 131 2 S56474 hypothetrical 14.6K

ALIGNMENTS

RESULT 1

F84162
hypothetical protein Vng0026c [imported] - Halobacterium sp. NRC-1
C;Species: Halobacterium sp. NRC-1
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C;Accession: F84162
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabc
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omar, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A;Title: Genome sequence of Halobacterium species NRC-1.
A;Reference number: A84160; MUID:20504483; PMID:11016950
A;Accession: F84162
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-307 <STO>
A;Cross-references: GI:AE004437; NID:G10579674; PIDN:AGI8666.1; GSPDB:GN00138
C;Genetics:
A;Gene: VNG0026C

Query Match 81.8%; Score 36; DB 2; Length 307;
Best Local Similarity 77.8%; Pred. No. 8.2;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 STAVFPFAGR 9

DB 92 STAVFPFGR 100
|||||
|||

RESULT 2

S32864
outJ protein - Erwinia carotovora
C;Species: Erwinia carotovora
C;Date: 08-Dec-1993 #sequence_revision 26-May-1995 #text_change 08-Oct-1999
C;Accession: S32864; S31753
R;Reeves, P.J.; Whitcombe, D.; Wharam, S.; Gibson, M.; Allison, G.; Bunce, N.; Barallón,
Mol. Microbiol. 8, 443-456, 1993
A;Title: Molecular cloning and characterization of 13 out genes from Erwinia carotovora
bacteria.
A;Reference number: S32857; MUID:93316842; PMID:8326859
A;Accession: S32864
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-216 <REE>
A;Cross-references: EMBL:X70049; NID:G42184; PIDN:CAA49651.1; PID:G42192
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, /./
C;Genetics:
A;Gene: outJ
C;Superfamily: secretion protein xcpW
Query Match 79.5%; Score 35; DB 2; Length 216;

```
; APPLICANT: Brooks, Peter
; APPLICANT: Cheres, David A.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS USEFUL FOR INHIBITION OF
; FILE REFERENCE: MER0049S
; CURRENT APPLICATION NUMBER: US/10/115,223
; PRIOR FILING DATE: 2002-04-02
; PRIOR APPLICATION NUMBER: US/09/194,468
; PRIOR FILING DATE: 1999-03-23
; PRIOR APPLICATION NUMBER: 60/018,773
; PRIOR FILING DATE: 1996-05-31
; PRIOR APPLICATION NUMBER: 60/015,896
; PRIOR FILING DATE: 1996-05-31
; PRIOR APPLICATION NUMBER: PCT/US97/09158
; PRIOR FILING DATE: 1997-05-30
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 26
; LENGTH: 108
; TYPE: PRT
; ORGANISM: Gallus gallus
US-10-115-223-26

Query Match      68.2%; Score 30; DB 12; Length 108;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3 AVFFAG 8
        |||||
DB      65 AVFFAG 70

Search completed: December 11, 2003, 18:38:05
Job time : 49.3333 secs
```

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR INHIBITING ANGIOGENESIS

```
; FILE REFERENCE: 07555.0001
; CURRENT APPLICATION NUMBER: US/09/414,834
; CURRENT FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 464
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Antithrombin III
US-09-414-834-1
```

Query Match 70.5%; Score 31; DB 9; Length 464;

Best Local Similarity 77.8%; Pred. No. 2.9e+02; Mismatches 0; Indels 0; Gaps 0;

Matches 7; Conservative 0;

QY 1 STAVFFAGR 9

Db 417 STAVVIAGR 425

RESULT 7

```
US-09-841-132-132
; Sequence 192, Application US/09841132
; Patent No. US20020061848A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Probst, Peter
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C8
; CURRENT APPLICATION NUMBER: US/09/841,132
; CURRENT FILING DATE: 2001-04-23
; NUMBER OF SEQ ID NOS: 599
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 192
; LENGTH: 848
; TYPE: PRT
; ORGANISM: Chlamydia
US-09-841-132-132
```

Query Match 70.5%; Score 31; DB 9; Length 848;

Best Local Similarity 55.6%; Pred. No. 5.5e+02;

Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 STAVFFAGR 9

Db 243 SDAIYFAGK 251

RESULT 8

```
US-09-841-132-178
; Sequence 178, Application US/09841132
; Patent No. US20020061848A1
; GENERAL INFORMATION:
; APPLICANT: Bhatia, Ajay
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Probst, Peter
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
; FILE REFERENCE: 210121.469C8
; CURRENT APPLICATION NUMBER: US/09/841,132
; CURRENT FILING DATE: 2001-04-23
; NUMBER OF SEQ ID NOS: 599
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 178
; LENGTH: 1530
; TYPE: PRT
; ORGANISM: Chlamydia
US-09-841-132-178
```

Query Match 70.5%; Score 31; DB 9; Length 1530;

Best Local Similarity 55.6%; Pred. No. 1e+03; Mismatches 1; Indels 0; Gaps 0;

Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 STAVFFAGR 9

Db 925 SDAIYFAGK 933

RESULT 11

```
US-10-156-761-10429
; Sequence 10429, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
```

Query Match 70.5%; Score 31; DB 14; Length 1531;

Best Local Similarity 55.6%; Pred. No. 1e+03; Mismatches 1; Indels 0; Gaps 0;

Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 STAVFFAGR 9

Db 925 SDAIYFAGK 933

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 18:21:27 ; Search time 49.3333 Seconds
(without alignments)
33.929 Million cell updates/sec

Title: US-10-014-658-15

Perfect score: 44

Sequence: 1 STAVFFAGR 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 684280 seqs, 185983659 residues

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	33	75.0	415	12 US-10-214-446-46	Sequence 46, Appli
2	32	72.7	35	12 US-10-355-780-7	Sequence 7, Appli
3	32	72.7	203	15 US-10-156-761-9311	Sequence 9111, Ap
4	31	70.5	56	14 US-10-043-452-13	Sequence 13, Appl
5	31	70.5	306	15 US-10-156-761-12838	Sequence 12838, A
6	31	70.5	464	9 US-09-414-834-1	Sequence 1, Appli
7	31	70.5	848	9 US-09-841-132-192	Sequence 192, App
8	31	70.5	1530	9 US-09-841-132-178	Sequence 178, App
9	31	70.5	1531	12 US-09-841-260-98	Sequence 98, Appl
10	31	70.5	1531	14 US-10-007-693-98	Sequence 98, Appl
11	31	70.5	3970	15 US-10-156-761-10429	Sequence 10429, A
12	30	68.2	74	12 US-10-115-223-19	Sequence 19, Appl
13	30	68.2	74	12 US-10-115-223-25	Sequence 25, Appl
14	30	68.2	108	12 US-10-115-223-20	Sequence 20, Appl
15	30	68.2	108	12 US-10-115-223-26	Sequence 26, Appl

16	30	68.2	119	9 US-09-726-899-7	Sequence 7, Appli
17	30	68.2	119	12 US-10-231-452-40	Sequence 40, Appl
18	30	68.2	119	12 US-10-231-452-41	Sequence 41, Appl
19	30	68.2	193	12 US-10-115-223-18	Sequence 18, Appl
20	30	68.2	222	12 US-10-115-223-17	Sequence 24, Appl
21	30	68.2	222	12 US-10-115-223-23	Sequence 17, Appl
22	30	68.2	228	12 US-10-115-223-23	Sequence 23, Appl
23	30	68.2	293	12 US-10-017-161-2306	Sequence 2306, Ap
24	30	68.2	311	9 US-09-883-720-16	Sequence 16, Appl
25	30	68.2	351	15 US-10-156-761-12617	Sequence 12617, A
26	30	68.2	412	9 US-09-795-693-21	Sequence 21, Appl
27	30	68.2	412	15 US-10-156-761-12617	Sequence 21, Appl
28	30	68.2	412	15 US-10-199-485-21	Sequence 21, Appl
29	30	68.2	429	12 US-10-115-223-45	Sequence 45, Appl
30	30	68.2	455	15 US-10-156-761-10488	Sequence 10488, A
31	30	68.2	482	15 US-10-156-761-11874	Sequence 11874, A
32	30	68.2	660	9 US-09-801-196-35	Sequence 35, Appl
33	30	68.2	660	12 US-10-153-185-14	Sequence 14, Appl
34	30	68.2	660	12 US-10-301-822-125	Sequence 125, App
35	30	68.2	660	12 US-10-219-561-14	Sequence 14, Appl
36	30	68.2	660	12 US-10-131-985-25	Sequence 25, Appl
37	30	68.2	660	15 US-10-219-329-14	Sequence 14, Appl
38	30	68.2	663	12 US-10-115-223-30	Sequence 30, Appl
39	30	68.2	2842	12 US-10-224-999A-3479	Sequence 3479, Ap
40	30	68.2	2905	8 US-08-424-550B-401	Sequence 401, App
41	29	65.9	34	12 US-10-355-780-6	Sequence 6, Appli
42	29	65.9	50	12 US-09-933-767-711	Sequence 711, App
43	29	65.9	50	12 US-10-023-282-711	Sequence 711, App
44	29	65.9	124	12 US-10-410-907A-32	Sequence 32, Appl
45	29	65.9	176	10 US-09-860-670-120	Sequence 120, App

ALIGNMENTS

RESULT 1

US-10-214-446-46
; Sequence 46, Application US/10214446
; Publication No. US20030180742A1
; GENERAL INFORMATION:
; APPLICANT: Weiner, David
; APPLICANT: Burk, Mark J.
; APPLICANT: Hitchman, Tim
; APPLICANT: Pujol, Catherine
; APPLICANT: Richardson, Toby
; APPLICANT: Short, Jay M.
; TITLE OF INVENTION: P450 ENZYMES, NUCLEIC ACIDS ENCODING
; TITLE OF INVENTION: THEM AND METHODS OF MAKING AND USING THEM
; FILE REFERENCE: 09010-500001
; CURRENT APPLICATION NUMBER: US/10/214,446
; CURRENT FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US 60/309,497
; PRIOR FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 46
; LENGTH: 415
; TYPE: PRT
; ORGANISM: Bacterial
US-10-214-446-46

Query Match 75.0%; Score 33; DB 12; Length 415;
Best Local Similarity 75.0%; Pred. No. 1e-02; 1; Indels 0; Gaps 0;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 TAVFFAGR 9

DB 81 TAAFFACK 88

RESULT 2

US-10-355-780-7
; Sequence 7, Application US/10355780

; TYPE: PRT
; ORGANISM: Chlamydia
US-09-620-412C-178

Query Match 70.5%; Score 31; DB 4; Length 1530;
Best Local Similarity 55.6%; Pred. No. 5.9e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 STAFFPAG 9
| | | | |
Db 925 SDIYFAGK 933

RESULT 15
US-09-598-419-178
; Sequence 178 Application US/09598419
; Patent No. 6565856
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir A.W.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND
; FILE REFERENCE: 210121.469C6
; CURRENT APPLICATION NUMBER: US/09/598,419
; CURRENT FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 357
; SOFTWARE: FastSeq for Windows Version 3.0/4.0
; SEQ ID NO 178
; LENGTH: 1530
; TYPE: PRT
; ORGANISM: Chlamydia
US-09-598-419-178

Query Match 70.5%; Score 31; DB 4; Length 1530;
Best Local Similarity 55.6%; Pred. No. 5.9e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 STAFFPAG 9
| | | | |
Db 925 SDIYFAGK 933

Search completed: December 11, 2003, 18:30:44
Job time : 10.2222 secs

NUMBER OF SEQUENCES: 552
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: Fast-SEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/858,207A
FILING DATE: 09-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/017670
FILING DATE: 14-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimmi, Edward R.
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P50475
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 273:
SEQUENCE CHARACTERISTICS:
LENGTH: 110 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: No. 6348328e
US-08-858-207A-273

Query Match 70.5%; Score 31; DB 4; Length 110;
Best Local Similarity 75.0%; Pred. No. 44;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TAVFFAGR 9
Db 28 TATFFIGR 35

RESULT 8
US-09-113-750A-40
Sequence 40, Application US/09113750A
Patent No. 6294176
GENERAL INFORMATION:
APPLICANT: David E. Junker and Mark D. Cochran
TITLE OF INVENTION: Recombinant Raccoonpox virus
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: John P. White
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/113,750A
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678

REFERENCE/DOCKET NUMBER: 55744
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)262-0400
TELEFAX: (212)664-0525
TELEX: 422523
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 200 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-113-750A-40

Query Match 70.5%; Score 31; DB 3; Length 200;
Best Local Similarity 77.8%; Pred. No. 80;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 STAVFFAGR 9
Db 97 STAGFFGR 105

RESULT 9
US-08-046-431A-2
Sequence 2, Application US/08046431A
Patent No. 5420252
GENERAL INFORMATION:
APPLICANT: KATO, HIROYUKI
APPLICANT: YOSHITAKE, SHINJI
APPLICANT: SUZUKI, SUGURA
APPLICANT: SUZUKI, NOBORU
APPLICANT: SETO, TOSHIO
APPLICANT: NAGAOKA, NAKO
APPLICANT: MIZUI, YOSHIHARU
TITLE OF INVENTION: HUMAN ANTITHROMBIN III MUTANT
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLYNN, THIEL, BOUTELL & TANIS, P.C.
STREET: 2026 Rambling Road
CITY: Kalamazoo
STATE: Michigan
COUNTRY: USA
ZIP: 49008-1699
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inches, 1.44 Mb storage
COMPUTER: IBM PC/XT/AT Compatible
OPERATING SYSTEM: MS-DOS 5.0
SOFTWARE: WordPerfect 5.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/046,431A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP90488
FILING DATE: 10-APR-1992
APPLICATION NUMBER: JP31855
FILING DATE: 22-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Terryence F. Chapman
REGISTRATION NUMBER: 32549
REFERENCE/DOCKET NUMBER: Furuya Case 1286
TELECOMMUNICATION INFORMATION:
TELEPHONE: (616) 381-1156
TELEFAX: (616) 381-5465
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 454 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
DESCRIPTION:
US-08-046-431A-2

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 18:21:01 ; Search time 8.22222 Seconds
(without alignments)
46.313 Million cell updates/sec

Title: US-10-014-658-15

Perfect score: 44

Sequence: 1 STAVFFAGR 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/2/iaa/5A COMB.pdp:*

2: /cgn2_6/ptodata/2/iaa/5B COMB.pdp:*

3: /cgn2_6/ptodata/2/iaa/6A COMB.pdp:*

4: /cgn2_6/ptodata/2/iaa/6B COMB.pdp:*

5: /cgn2_6/ptodata/2/iaa/PCTUS COMB.pdp:*

6: /cgn2_6/ptodata/2/iaa/backfiles1.pdp:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	33	75.0	452	US-09-328-352-7869	Sequence 7869, Ap
2	32	72.7	35	US-09-627-2183-7	Sequence 7, Appl
3	31	70.5	56	US-09-271-608-13	Sequence 13, Appl
4	31	70.5	56	US-09-695-950-13	Sequence 13, Appl
5	31	70.5	56	US-09-696-147-13	Sequence 13, Appl
6	31	70.5	56	US-09-696-364-13	Sequence 13, Appl
7	31	70.5	110	US-08-858-207A-273	Sequence 273, App
8	31	70.5	200	US-09-113-750A-40	Sequence 40, Appl
9	31	70.5	464	US-08-046-431A-2	Sequence 2, Appl
10	31	70.5	848	US-09-556-877-192	Sequence 192, App
11	31	70.5	848	US-09-620-412C-192	Sequence 192, App
12	31	70.5	848	US-09-598-419-192	Sequence 192, App
13	31	70.5	1530	US-09-556-877-178	Sequence 178, App
14	31	70.5	1530	US-09-620-412C-178	Sequence 178, App
15	31	70.5	1530	US-09-598-419-178	Sequence 178, App
16	30	68.2	74	US-09-194-468A-19	Sequence 19, Appl
17	30	68.2	74	US-09-194-468A-25	Sequence 25, Appl
18	30	68.2	108	US-09-194-468A-20	Sequence 20, Appl
19	30	68.2	108	US-09-194-468A-26	Sequence 26, Appl
20	30	68.2	113	US-07-789-344A-10	Sequence 10, Appl
21	30	68.2	115	US-08-428-197-42	Sequence 42, Appl
22	30	68.2	115	PCT-US93-10555-42	Sequence 42, Appl
23	30	68.2	119	US-08-785-065-7	Sequence 7, Appl
24	30	68.2	119	US-09-151-412-7	Sequence 7, Appl
25	30	68.2	193	US-09-194-468A-18	Sequence 18, Appl
26	30	68.2	193	US-09-194-468A-24	Sequence 24, Appl
27	30	68.2	222	US-09-194-468A-17	Sequence 17, Appl

28 68.2 228 4 US-09-194-468A-23 Sequence 23, Appl
29 30 68.2 311 3 US-09-282-305-16 Sequence 16, Appl
30 30 68.2 311 4 US-09-883-720-16 Sequence 16, Appl
31 30 68.2 421 4 US-09-107-532A-4030 Sequence 4030, Ap
32 30 68.2 429 4 US-09-194-468A-45 Sequence 45, Appl
33 30 68.2 631 3 US-08-448-489-17 Sequence 17, Appl
34 30 68.2 660 3 US-08-704-711A-18 Sequence 18, Appl
35 30 68.2 660 4 US-09-521-220-18 Sequence 19, Appl
36 30 68.2 660 4 US-09-391-104-19 Sequence 30, Appl
37 30 68.2 663 4 US-09-194-468A-30 Sequence 5, Appl
38 30 68.2 1507 3 US-08-929-329-5 Sequence 401, App
39 30 68.2 2305 4 US-08-469-260A-401 Sequence 401, App
40 30 68.2 2305 4 US-08-488-446-401 Sequence 401, App
41 30 68.2 2305 4 US-08-467-344A-401 Sequence 401, App
42 29 65.9 34 4 US-09-627-218B-6 Sequence 6, Appl
43 29 65.9 50 4 US-09-205-258-711 Sequence 711, App
44 29 65.9 250 1 US-08-133-804-2 Sequence 2, Appl
45 29 65.9 250 1 US-08-461-184-8 Sequence 8, Appl

ALIGNMENTS

RESULT 1
US-09-328-352-7869
; Sequence 7869, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7869
; LENGTH: 452
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-7869

Query Match 75.0%; Score 33; DB 4; Length 452;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 STAVFFA 7
DB 326 STAVFFA 332

RESULT 2
US-09-627-218B-7
; Sequence 7, Application US/09627218B
; Patent No. 6537548
; GENERAL INFORMATION:
; APPLICANT: Prusiner, Stanley
; APPLICANT: Safar, Jiri
; APPLICANT: Williamson, Anthony
; APPLICANT: Burton, Dennis
; TITLE OF INVENTION: Antibodies Specific for Ungulate Prp
; FILE REFERENCE: UCAL-194
; CURRENT APPLICATION NUMBER: US/09/627,218B
; CURRENT FILING DATE: 2000-07-27
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 35
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: synthesized peptide
US-09-627-218B-7

CC ABS56454. Also included are an antibody which binds one of the
 CC proteins, treating a patient by administering the protein, DNA or
 CC antibody (in a composition), a kit comprising first and second primers,
 CC which are the nucleic acid cited above or fragments between nucleotides
 CC 8-100 of a sequence not defined in the specification, for amplifying a
 CC target sequence contained within a Streptococcus nucleic acid sequence,
 CC where the first primer is substantially complementary to the target
 CC sequence and the second primer is substantially complementary to the
 CC complement of the target sequence, and where the parts of the primers
 CC having substantial complementarity define the termini of the target
 CC sequence to be amplified, assay comprising contacting a test compound
 CC with the protein, and determining whether the test compound binds to the
 CC protein and a Streptococcus pneumoniae bacterium, where one or more
 CC genes encoding the proteins has been rendered inactive. The proteins,
 CC nucleic acid molecules, antibody and compositions are useful as
 CC medicaments for treating or preventing a disease or infection due to
 CC streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,
 CC sepsis, otitis media or ear infection. They are also useful in developing
 CC vaccines, diagnostics and antibiotics. The methods are useful for
 CC identifying immunodominant proteins. The present sequence is one of
 CC the 2469 proteins expressed by the identified coding regions from the
 CC genomic sequence.

CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX
 SQ Sequence 115 AA;
 Query Match 70.5%; Score 31; DB 24; Length 115;
 Best Local Similarity 75.0%; Pred. No. 2.2e+02;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TAVFFAGR 9
 |||||
 Db 28 TATFFIGR 35

RESULT 15

AAV85965
 ID AAY85965 standard; Protein: 122 AA.

XX AC AAY85965;

XX DT 10-APR-2000 (first entry)

XX S. pneumoniae derived protein #174.

XX Treatment; prevention; disease; diagnosis; gene therapy; screening;
 KW bacterial; antimicrobial; antibiotic; pathogenesis; infection.

XX Streptococcus pneumoniae.

XX PN WO9806734-A1.

XX PD 19-FEB-1998.

XX PF 15-AUG-1997; 97WO-US14436.

XX PR 16-AUG-1996; 96US-0024022.

XX PA (SMIK) SMITHKLINE BEECHAM CORP.

XX PI Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;
 PI Stodola RK;

XX DR WPI; 1998-159452/14.

XX DR N-PSDB; AAZ96316.

XX Streptococcus pneumoniae proteins and related DNA - useful for
 PT screening compounds for antibacterial activity

XX Claim 5; Page 460; 640pp; English.

XX
 CC This invention describes novel isolated Streptococcus pneumoniae
 CC polynucleotides (see AAZ96173-296494) and their encoded proteins (see
 CC AAY85792-Y86182). The DNA, vectors and host cells described in the
 CC method of the invention are useful for the recombinant expression of the
 CC polypeptides. The polypeptides are useful for treatment or prevention of
 CC disease, or diagnosis of disease related to expression or activity of
 CC such a polypeptide. They can also be used to screen for compounds which
 CC interact with and inhibit or activate such a polypeptide. The
 CC polypeptides (or DNA encoding them, via gene therapy) are also useful
 CC for inducing an immunological response in a mammal. The antagonists are
 CC useful to inhibit such bacterial polypeptides. The polypeptides are
 CC particularly useful to identify antimicrobial compounds and antibiotics.
 CC They are also useful to determine their role in pathogenesis of
 CC infection, dysfunction and disease.

XX SQ Sequence 122 AA;

Query Match 70.5%; Score 31; DB 19; Length 122;
 Best Local Similarity 75.0%; Pred. No. 2.3e+02;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TAVFFAGR 9
 |||||
 Db 28 TATFFIGR 35

Search completed: December 11, 2003, 18:25:04
 Job time : 27 secs

```

RN SEQUENCE FROM N.A.
RC STRAIN=Bristol N2; PubMed=7906398;
RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kerhaw J., Kirsten J., Laister N.,
RA Lattelle P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roop A., Saunders D., Shownkeen R.,
RA Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Worldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP REVISIONS.
RA Waterston R.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L10986; AAA28013.2; -.
DR WormPep; F10E9.2; CE24893.
KW Hypothetical protein.
SQ SEQUENCE 286 AA; 31609 MW; 61751F182939DD57 CRC64;

Query Match 76.1%; Score 35; DB 1; Length 286;
Best Local Similarity 81.8%; Fred. No. 31;
Matches 9; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

QY 1 SEEGEG--SGR 9
DB 202 SEEGEGFSGR 212
|||||
ID HTPX_VIBCH STANDARD; PRT; 287 AA.
AC Q9K5Y9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable protease htpx homolog (EC 3.4.24.-).
GN HTPX OR VC117.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=566;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=E1 Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bess S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Niernan W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RL Nature 406:477-483(2000).
CC -!- COFACTOR: Binds 1 zinc ion per subunit (By similarity).

```

```

CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M48.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB004192; AAF94276.1; -.
DR FIR; E82239; E82239.
DR MEROPS; M48.002; -.
DR TIGR; VC1117; -.
DR HAMAP; MF_00188; -.
DR InterPro; IPR001915; Peptidase_M48.
DR InterPro; IPR006025; Zn_MTPeptide.
DR Pfam; PF01435; Peptidase_M48; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; Zinc; Complete proteome.
KW Transmembrane; Hydrolase; Metalloprotease; Zinc; Complete proteome.
FT TRANSMEM 7 29 POTENTIAL.
FT TRANSMEM 34 56 POTENTIAL.
FT TRANSMEM 162 179 POTENTIAL.
FT TRANSMEM 194 216 POTENTIAL.
FT ACT_SITE 144 144 BY SIMILARITY.
FT ACT_SITE 143 143 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 147 147 ZINC (CATALYTIC) (BY SIMILARITY).
SQ SEQUENCE 287 AA; 31229 MW; 3EA87D30B64E9EAB CRC64;

Query Match 76.1%; Score 35; DB 1; Length 287;
Best Local Similarity 100.0%; Fred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SEEGEGS 7
DB 183 SEEGEGS 189
|||||
ID TYSY_YEAST STANDARD; PRT; 304 AA.
AC P06785; Q12694;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Thymidylate synthase (EC 2.1.1.45) (TS) (TSase).
GN TMF1 OR CDC21 OR YOR074C OR YOR29-25.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87165970; PubMed=3031048;
RA Taylor G.R., Lagosky P.A., Storms R.K., Haynes R.H.;
RT "Molecular characterization of the cell cycle-regulated thymidylate
RT synthase gene of Saccharomyces cerevisiae.";
RL J. Biol. Chem. 262:5298-5307(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97279235; PubMed=9133743;
RA Valens M., Bohn C., Daigian-Fornier B., Dang V., Bolotin-Fukuhara M.;
RT "The sequence of a 54.7 kb fragment of yeast chromosome XV reveals
RT the presence of two tRNAs and 24 new open reading frames.";
RL Yeast 13:379-390(1997).
RN [3]
RP SEQUENCE OF 1-38 FROM N.A.
RX MEDLINE=89096830; PubMed=3062362;
RA McIntosh E.M., Ord R.W., Storms R.K.;
RT "Transcriptional regulation of the cell cycle-dependent thymidylate
RT synthase gene of Saccharomyces cerevisiae.";
RL Mol. Cell. Biol. 8:4616-4624(1988).

```



```
FT CONFLICT 337 337 D -> G (IN REF. 3).
FT CONFLICT 446 446 L -> W (IN REF. 3).
FT CONFLICT 584 584 S -> G (IN REF. 3).
SQ SEQUENCE 1189 AA; 127509 MW; 5E244858716E5DF CRC64;

Query Match 76.1%; Score 35; DB 1; Length 1189;
Best Local Similarity 87.5%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SEEGEGSG 8
|||
|||
|||
DB 538 SEEGPGSG 545

RESULT 11
ADDA_BACSU
ID ADDA_BACSU STANDARD; PRT; 1232 AA.
AC P23478;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE ATP-dependent nuclease subunit A.
GN ADDA.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]_TaxID=1423;
RP SEQUENCE FROM N.A.
RC STRAIN=OG1;
RX MEDLINE=91267926; PubMed=1646786;
RA Kozlstra J., Venema G.;
RT "Cloning, sequencing, and expression of Bacillus subtilis genes
involved in ATP-dependent nuclease synthesis.";
RL J. Bacteriol. 173:3644-3655(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98240224; PubMed=9579061;
RA Noback M.A., Holsappel S., Kiewiet R., Terpstra P., Wambutt R.,
Wiedler H., Venema G., Bron S.;
RT "The 172 kb prkA-addA region from 83 degrees to 97 degrees of the
Bacillus subtilis chromosome contains several dysfunctional genes,
the glyA marker, many genes encoding transporter proteins, and the
ubiquitous hit gene.";
RL Microbiology 144:859-875(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
Ghim S.-Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
Kurita K., Lapidus A., Lardinis S., Lauber J., Lazarevic V.,
Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
Sato T., Scanlan E., Schleich S., Schroter R., Scoffone F.,
Sekiguchi J., Sekowska A., Seror S.J., Serrero P., Shin B.S., Soldo B.,
Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,
Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
Viati A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
subtilis.";
RL Nature 390:249-256(1997).
RN [4]
RP SEQUENCE OF 1000-1232 FROM N.A.
RC STRAIN=168;
RX MEDLINE=98015415; PubMed=9359311;
RA Medina N., Vannier F., Roche B., Autret S., Levine A., Seror S.J.;
RT "Sequencing of regions downstream of addA (98 degrees) and citG (289
degrees) in Bacillus subtilis.";
RL Microbiology 143:3305-3308(1997).
CC -!- FUNCTION: THE ENZYME COMPLEX MAY HAVE A WIDE VARIETY OF CATALYTIC
ACTIVITIES INCLUDING ATP-DEPENDENT EXONUCLEASE, ATP-STIMULATED
ENDONUCLEASE, ATP-DEPENDENT HELICASE AND DNA-DEPENDENT ATPASE
ACTIVITIES. THE SUBUNIT A MAY HAVE A ROLE IN DNA UNWINDING
(HELICASE), AND IN THE NUCLEASE ACTIVITY (BY SIMILARITY).
CC -!- SUBUNIT: THE B.SUBTILIS ATP-DEPENDENT NUCLEASE COMPLEX IS FORMED
BY ONE B SUBUNIT IN CONJUNCTION WITH ONE A SUBUNIT.
CC -!- SIMILARITY: IN 8 REGIONS TO E.COLI RECB PROTEIN, AND IN 6
REGIONS TO E.COLI UVRD AND REP HELICASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M63489; AA22201.1; -
CC EMBL; Y14081; CAA74482.1; -
CC EMBL; Z99109; CAB12903.1; -
CC EMBL; Y09476; CAA70668.1; -
CC PIR; B39432; B39432.
CC HSSP; P56255; 1PJR.
CC Subtilist; BG10466; addA.
CC InterPro; IPR000312; UvrD-helicase.
CC Pfam; PF00580; UvrD-helicase; 1.
KW Hydrolase; Nuclease; Exonuclease; Endonuclease; Helicase; ATP-binding;
KW DNA repair; Complete proteome.
FT NP BIND 30 37 ATP (POTENTIAL).
SQ SEQUENCE 1232 AA; 141072 MW; B8A115B16C3D4163 CRC64;

Query Match 76.1%; Score 35; DB 1; Length 1232;
Best Local Similarity 66.7%; Pred. No. 1.5e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 SEEGEGSGR 9
:|:|:|:|
DB 462 TEEGEGTGR 470

RESULT 12
HIC2_MOUSE
ID HIC2_MOUSE STANDARD; PRT; 266 AA.
AC Q9JLZ6;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hypermethylated in cancer 2 protein (Hic-2) (Fragment).
GN HIC2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Swiss Webster; TISSUE=Embryo;
RA Grimm C., Graw J.;
```

RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White C.J., Alonso J., Altschul H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Creasy T.H., Dewar K.,
RA Dunn P., Egu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Krenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
RA Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Utterback T., Van Aken S., Vaynsberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Xu G., Frazer C.M., Venter J.C., Davis R.W.,
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT thaliana";
RL Nature 408:816-820(2000).
CC -!- FUNCTION: ESSENTIAL FOR PROPER DEVELOPMENT OF LEAVES AND FLORAL
CC ORGANS, AND FORMATION OF AXILLARY MERISTEMS.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -!- TISSUE SPECIFICITY: WIDELY EXPRESSED AT LOW LEVELS.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT ALL DEVELOPMENTAL
CC STAGES.
CC -!- SIMILARITY: BELONGS TO THE ARGONAUTE FAMILY.
CC -!- SIMILARITY: Contains 1 PAZ domain.
CC -!- SIMILARITY: Contains 1 Piwi domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U91995; AAC18440.1; -
CC EMBL: ACC07932; AAD49755.1; -
CC InterPro: IPR003100; PAZ.
CC InterPro: IPR003165; Piwi.
CC Pfam: PF02170; PAZ; 1.
CC Pfam: PF02171; Piwi; 1.
CC PROSITE: PS50821; PAZ; 1.
CC PROSITE: PS50822; PIWI; 1.
KW Developmental protein.
FT DOMAIN 391 501 PAZ.
FT DOMAIN 676 997 PIWI.
FT DOMAIN 13 104 GLY-RICH.
SQ SEQUENCE 1048 AA; 116190 MW; 3E5146343A09C541 CRC64;
Query Match 73.9%; Score 34; DB 1; Length 1048;
Best Local Similarity 87.5%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 SEEGEGSG 8
Db 11 SEEGEGSG 18
RESULT 15
SMC3_RAT
ID SMC3_RAT STANDARD; PRT; 1191 AA.
AC P97690;
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Structural maintenance of chromosome 3 (chondroitin sulfate
DE proteoglycan 6) (Chromosome segregation protein SMC3) (Bamacan)
DE (Basement membrane-associated chondroitin proteoglycan).
GN CSPG6 OR SMC3 OR SMC3L1 OR BAM OR BMT.
OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Volk;
RX MEDLINE=97167695; PubMed=9015313;
RA Wu R.-R., Couchman J.R.;
RT "cDNA cloning of the basement membrane chondroitin sulfate
RT proteoglycan core protein, bamacan: a five domain structure including
RT coiled-coil motifs";
RL J. Cell Biol. 136:433-444(1997).
RN [2]
RP INTERACTION WITH SYCP2.
RX MEDLINE=20119364; PubMed=10652260;
RA Eljpe M., Heyting C., Gross B., Jessberger R.;
RT "Association of mammalian SMC1 and SMC3 proteins with meiotic
RT chromosomes and synaptonemal complexes";
RL J. Cell Sci. 113:673-682(2000).
CC -!- FUNCTION: Involved in chromosome cohesion during cell cycle and in
CC DNA repair. Central component of cohesin complex. The cohesin
CC complex is required for the cohesion of sister chromatids after
CC DNA replication. The cohesin complex apparently forms a large
CC proteinaceous ring within which sister chromatids can be trapped.
CC At anaphase, the complex is cleaved and dissociates from
CC chromatin, allowing sister chromatids to segregate. The cohesin
CC complex may also play a role in spindle pole assembly during
CC mitosis (By similarity).
CC -!- SUBUNIT: Forms a heterodimer with SMC1L1 or SMC1L2 in cohesin
CC complexes. Cohesin complexes are composed of the SMC1 (SMC1L1 or
CC SMC1L2) and SMC3 heterodimer attached via their hinge domain,
CC RAD21 which link them, and one STAG protein (STAG1, STAG2 or
CC STAG3), which interacts with RAD21. Also found in meiosis-specific
CC cohesin complexes. Interacts with NUPA1, and forms a ternary
CC complex with KIF3B and KIFAP3, suggesting a function in tethering
CC the chromosomes to the spindle pole and a function in chromosome
CC movement. Interacts with MXI1, MXD3 and MXD4 (By similarity).
CC Interacts with SYCP2.
CC -!- SUBCELLULAR LOCATION: Nuclear protein. Associates with chromatin.
CC Before prophase it is scattered along chromosome arms. During
CC prophase, most of cohesin complexes dissociate from chromatin
CC probably because of phosphorylation by PLK, except at centromeres,
CC where cohesin complexes remain. At anaphase, the RAD21 subunit of
CC the cohesin complex is cleaved, leading to the dissociation of the
CC complex from chromosomes, allowing chromosome separation (By
CC similarity).
CC -!- TISSUE SPECIFICITY: Ubiquitous.
CC -!- DOMAIN: The flexible hinge domain, which separates the large
CC intramolecular coiled coil regions, allows the heterotypic
CC interaction with the corresponding domain of SMC1L1 or SMC1L2,
CC forming a V-shaped heterodimer. The two heads of the heterodimer
CC are then connected by different ends of the cleavable RAD21
CC protein, forming a ring structure (By similarity).
CC -!- SIMILARITY: Belongs to the SMC family, SMC3 subfamily.
CC -!- CAUTION: Was originally isolated (Ref.1) as a proteoglycan protein
CC (explaining its name). Although not excluded, such secreted
CC function is not clear.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U82626; AAB96342.1; -
CC InterPro: IPR003405; SMC_C.
CC InterPro: IPR003395; SMC_N.
CC Pfam: PF02483; SMC_C; 1.
CC Pfam: PF02463; SMC_N; 1.
CC Mitosis; Meiosis; Cell cycle; Chromosome partition; DNA repair;
KW ATP-binding; Coiled coil; Nuclear protein.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 18:03:30 ; Search time 20.3333, Seconds
(without alignments)
114.220 Million cell updates/sec

Title: US-10-014-658-8

Perfect score: 46

Sequence: 1 SEEGEGSGR 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 23.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39	84.8	128	5 Q81Q40	Q81q40 drosophila
2	39	84.8	284	10 Q3FTW2	Q3ftw2 oryza sativ
3	38	82.6	509	10 Q40870	Q40870 picea glauc
4	38	82.6	820	16 Q8RAM1	Q8ram1 thermococ
5	38	82.6	1851	5 Q8T5G8	Q8t5g8 meloidogyne
6	37	80.4	131	16 Q8EPY8	Q8epy8 oceanobacil
7	37	80.4	168	11 Q923Z4	Q923z4 mus musculu
8	37	80.4	447	10 Q9M4H1	Q9m4h1 vitis vinif
9	37	80.4	711	10 Q8S3S1	Q8s3s1 oryza sativ
10	37	80.4	731	10 Q8H922	Q8h922 oryza sativ
11	37	80.4	1023	10 Q9LWH6	Q9lwh6 oryza sativ
12	37	80.4	1499	5 Q9U2M1	Q9u2m1 caenorhabdi
13	36	78.3	251	5 Q8MYW3	Q8myw3 drosophila
14	36	78.3	260	12 Q89238	Q89238 wheat dwarf
15	36	78.3	260	12 Q8UYL1	Q8uy11 wheat dwarf
16	36	78.3	295	4 Q9H746	Q9h746 homo sapien

ALIGNMENTS

RESULT 1

Q81Q40 PRELIMINARY; PRT; 128 AA.
ID Q81Q40
AC Q81Q40; (T-REMBLrel. 23, Created)
DT 01-MAR-2003 (T-REMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (T-REMBLrel. 23, Last sequence update)
DE CG32510-PA.
GN CG32510.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Spheroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Anantides P.G., Scher S.E., Li P.W., Hostins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.A.,
RA Brandon R.C., Rogers Y.H., Blaise R.G., Champs M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Agayani A., An H.J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Buchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Durkin B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foeller C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.D., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Housh K.A., Howland T.J., Wei M.H., Ibegwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

Q8full corynebacte
Q9yk86 peanut stri
Q8twa1 methanopyru
Q948h6 oryza sativ
Q8zpp8 salmonella
Q8z6k3 salmonella
P74848 salmonella
Q9vkc3 drosophila
Q8men9 drosophila
Q9x86 homo sapien
Q9hau3 homo sapien
Q9hbb5 homo sapien
Q9v629 homo sapien
Q8zn14 salmonella
Q8z5a0 salmonella
Q8ifb0 trypanosoma
Q9th02 homo sapien
Q8wv48 homo sapien
Q8lb27 arabidopsis
Q93wg3 arabidopsis
Q8hl67 arabidopsis
Q9frel arabidopsis
Q8ar36 mus musculu
Q93my3 streptococc
Q8dte4 streptococc
Q75043 homo sapien
Q8n3f1 homo sapien
Q16883 caenorhabdi


```
RC STRAIN=MB4 / JCM 11007;
RX MEDLINE=21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.,
RT "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700(2002).
DR EMBL; AE013082; AAM24422.1; -.
DR InterPro; IPR003156; DHAL.
DR InterPro; IPR001667; Ppsterase.
DR InterPro; IPR004610; RecJ.
DR Pfam; PF01368; DHX; 1.
DR Pfam; PF02272; DHAL; 1.
DR TIGRFAMs; TIGR00644; recJ; 1.
KW Exonuclease; Complete proteome.
SQ SEQUENCE 820 AA; 93109 MW; 497768C93914D6 CRC64;

Query Match 82.6%; Score 38; DB 16; Length 820;
Best Local Similarity 87.5%; Pred. No. 1.3e+02;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 BEGEGSGR 9
DB 377 BEGEGSGR 384

RESULT 5
Q8TSG8 PRELIMINARY; PRT; 1851 AA.
AC Q8TSG8;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Chitin synthase.
OS Meloidogyne artillia.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
OC Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.
OX NCBI_TaxID=42426;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21473256; PubMed=11589574;
RA Veronico P., Gray L.J., Jones J.T., Bazzicalupo P., Arbucci S.,
RA Cortese M.R., Di Vito M., De Giorgi C.;
RT "Nematode chitin synthases: Gene structure, expression and function in
RT Caenorhabditis elegans and the plant parasitic nematode Meloidogyne
RT artillia."
RL Mol. Genet. Genomics 266:28-34(2001).
DR EMBL; AV013285; AAG40111.1; -.
DR InterPro; IPR004835; Fungi_chitin_syn.
DR Pfam; PF03142; Chitin_synth_2; 1._syn.
SQ SEQUENCE 1851 AA; 210666 MW; 4B63A6DF9505E416 CRC64;

Query Match 82.6%; Score 38; DB 5; Length 1851;
Best Local Similarity 87.5%; Pred. No. 3.2e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 BEGEGSGR 9
DB 1777 BEGEGSGR 1784

RESULT 6
Q8EPY8 PRELIMINARY; PRT; 131 AA.
AC Q8EPY8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Cytochrome c551.
GN OB1942.
OS Oceanobacillus iheyensis.
OC Bacteria; Firmicutes; Bacillales; Oceanobacillus.
OX NCBI_TaxID=182710;
```

```
RN SEQUENCE FROM N.A.
RC STRAIN=TE831 / DSM 14371 / JCM 11309;
RX MEDLINE=22220767; PubMed=12235376;
RA Takami H., Takaki Y., Uchiyama I.;
RT "Genome sequence of Oceanobacillus iheyensis isolated from the Iheya
RT Ridge and its unexpected adaptive capabilities to extreme
RT environments."
RL Nucleic Acids Res. 30:3927-3935(2002).
DR EMBL; AF004599; BAC13898.1; -.
KW Complete proteome.
SQ SEQUENCE 131 AA; 13702 MW; 9AB282FE503C40F7 CRC64;

Query Match 80.4%; Score 37; DB 16; Length 131;
Best Local Similarity 87.5%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SEEEGSGR 8
DB 48 SEEEGSGR 55

RESULT 7
Q923Z4 PRELIMINARY; PRT; 168 AA.
AC Q923Z4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Nephew of atonal 3 (N-twist basic helix-loop-helix protein).
GN FERD3L OR NATO3 OR NTWIST.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=129/SV;
RX MEDLINE=21366020; PubMed=11472856;
RA Seggev E., Halachmi N., Salzberg A., Ben-Arie N.;
RT "Nato3 is an evolutionarily-conserved bHLH transcription factor
RT expressed in the CNS of Drosophila and mouse."
RL Mech. Dev. 106:197-202(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC Black B.L., Verzi M.P., Anderson J.P.;
RT "N-Twist, an evolutionarily-conserved bHLH protein expressed in the
RT developing CNS, functions as a transcriptional inhibitor."
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF369896; AAK72955.1; -.
DR MGD; MGI:2150010; Ferd3l.
DR InterPro; IPR001092; HLH_basic.
DR Pfam; PF00010; HLH; 1.
DR PROSITE; PS00038; HLH_1; 1.
DR PROSITE; PS50888; HLH_2; 1.
SQ SEQUENCE 168 AA; 19463 MW; 960BA60C19E7237B CRC64;

Query Match 80.4%; Score 37; DB 11; Length 168;
Best Local Similarity 87.5%; Pred. No. 35;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 BEGEGSGR 9
DB 79 BEGEGSGR 86

RESULT 8
Q9M4H1 PRELIMINARY; PRT; 447 AA.
ID Q9M4H1
AC Q9M4H1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
```

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 SEEGEGSGR 9

Db 456 SEEGHGDGR 464

RESULT 12

Q9U2M1 PRELIMINARY; PRT; 1499 AA.

AC Q9U2M1, 2000 (Tremblrel. 13, Created)
DT 01-MAY-2002 (Tremblrel. 22, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 23, Last annotation update)
DE Y39A1B.3 protein.
GN Y39A1B.3
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Wall M.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95069613; PubMed=9851916;
RA none;
RT "Genome sequencing of the nematode C.elegans: A platform for
investigating biology."
RL Science 282:2012-2018 (1998).
DR EMBL; AL021482; CAAL6340.3; -
DR WormPep; Y39A1B.3; CE31734;
SQ SEQUENCE 1499 AA; 171021 MW; 779D91C231668355 CRC64;

Query Match 80.4%; Score 37; DB 5; Length 1499;
Best Local Similarity 87.5%; Pred. No. 3.9e+02;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 EEEGSGR 9

Db 321 ENGEGSGR 328

RESULT 13

Q8MYW3 PRELIMINARY; PRT; 251 AA.

AC Q8MYW3
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE RH01011p.
GN CG14945.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=Berkley;
RA Stapleton M., Brokstein P., Hong L., Aghayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George K., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunoo J., Pacleab J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AV113552; AM29557.1; -
DR Flybase; FBgn032402; CG14945.
DR InterPro; IPR003633; Varsuifglyc_PPLC.
DR Pfam; PF03490; Varsuif_PPLC; 1.
SQ SEQUENCE 251 AA; 27650 MW; 18F35A5E5F0B518C CRC64;

Query Match 78.3%; Score 36; DB 5; Length 251;
Best Local Similarity 87.5%; Pred. No. 84;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SEEGEGSG 8

Db 89 SEEGSGSG 96

RESULT 14

Q89238 PRELIMINARY; PRT; 260 AA.
AC Q89238
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Capsid protein V2.
GN V2.
OS Wheat dwarf virus (WDV).
OC Viruses; ssDNA viruses; Geminiviridae; Mastrevirus.
OX NCBI_TaxID=10834;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=French;
RA Bendahmane M., Schalk H.J., Gronenborn B.;
RT "Identification and characterization of wheat dwarf virus (WDV) from
France."
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; X82104; CAA57624.1; -
DR InterPro; IPR00143; Gencoat_MSV.
DR Pfam; PF00844; Gemini_coat; 1-
DR PRINTS; PR00223; GEMCOATABR1.
DR ProDom; PD001964; Gencoat_MSV; 1.
SQ SEQUENCE 260 AA; 29363 MW; 05B3A61A95F77BEB CRC64;

Query Match 78.3%; Score 36; DB 12; Length 260;
Best Local Similarity 87.5%; Pred. No. 88;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 EEEGSGR 9

Db 16 EEEGSGR 23

RESULT 15

Q8UYL1 PRELIMINARY; PRT; 260 AA.

AC Q8UYL1
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Coat protein.
GN V2.
OS Wheat dwarf virus (WDV).
OC Viruses; ssDNA viruses; Geminiviridae; Mastrevirus.
OX NCBI_TaxID=10834;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=Wheat;
RA Kvarnheden A.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=Wheat;
RC MEDLINE=21844413; PubMed=11858231;
RA Kvarnheden A., Lindblad M., Lindsten K., Valkonen J.P.T.;
RT "Genetic diversity of Wheat dwarf virus."
RL Arch. Virol. 147:205-216(2002).
DR EMBL; AJ311031; CAC84659.1; -
DR InterPro; IPR000143; Gencoat_MSV.
DR Pfam; PF000263; GV_A/BRI_coat.
SQ SEQUENCE 260 AA; 29363 MW; 05B3A61A95F77BEB CRC64;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 17:22:25 ; Search time 25 Seconds
(without alignments)
57.142 Million cell updates/sec

Title: US-10-014-658-10
Perfect score: 44
Sequence: 1 STEVEGAGR 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq_19Jun03.*

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	44	100.0	9	21	AA1980
2	41	93.2	9	21	AA1981
3	38	86.4	9	21	AA1982
4	38	86.4	9	21	AA1983
5	36	81.8	359	21	AA1988
6	36	81.8	399	21	AA1989
7	35	79.5	9	21	AA1990
8	35	79.5	529	21	AA1991
9	34	77.3	9	21	AA1992

10	34	77.3	135	22	AA1990
11	34	77.3	730	23	AA1991
12	33	75.0	294	22	AA1992
13	33	75.0	514	22	AA1993
14	33	75.0	2858	22	AA1994
15	33	75.0	3060	22	AA1995
16	32	72.7	9	21	AA1996
17	32	72.7	80	22	AA1997
18	32	72.7	82	22	AA1998
19	32	72.7	100	22	AA1999
20	32	72.7	181	22	AA2000
21	32	72.7	275	22	AA2001
22	32	72.7	456	23	AA2002
23	32	72.7	586	22	AA2003
24	32	72.7	1167	15	AA2004
25	32	72.7	1167	18	AA2005
26	32	72.7	1167	18	AA2006
27	32	72.7	1167	20	AA2007
28	32	72.7	1167	20	AA2008
29	31	70.5	42	22	AA2009
30	31	70.5	54	22	AA2010
31	31	70.5	54	22	AA2011
32	31	70.5	54	22	AA2012
33	31	70.5	54	22	AA2013
34	31	70.5	54	22	AA2014
35	31	70.5	54	22	AA2015
36	31	70.5	54	22	AA2016
37	31	70.5	54	22	AA2017
38	31	70.5	54	22	AA2018
39	31	70.5	54	23	AA2019
40	31	70.5	57	23	AA2020
41	31	70.5	88	23	AA2021
42	31	70.5	127	21	AA2022
43	31	70.5	127	22	AA2023
44	31	70.5	169	19	AA2024
45	31	70.5	178	23	AA2025

ALIGNMENTS

RESULT 1
AA1980
ID AAY44474 standard; peptide; 9 AA.
XX AAY44474;
XX AAY44474;
DT 27-MAR-2000 (first entry)
XX Human antithrombin III variant 13.B (residues 385-393).

Human; antithrombin III; ATIII variant 13.B; elastase-resistant;
KW IGG activated neutrophil resistant; anti-thrombin activity; heparin;
KW anti-factor Xa activity; blood clotting disorder; sepsis; trauma; stroke;
KW thrombin activation-related pathological symptom; restenosis; thrombosis;
KW acute respiratory distress syndrome; thromboembolism; reocclusion.

XX Homo sapiens.
OS Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 3 /note= "ATIII.N135A Ala at 387 is substituted by Glu"

FT Misc-difference 5 /note= "ATIII.N135A Val at 389 is substituted by Glu"

FT Misc-difference 6 /note= "ATIII.N135A Ile at 390 is substituted by Gly"

PN WO958098-A2.

PD 18-NOV-1999.

XX 12-MAY-1999; 99WO-US10549.

C glutamicum prote
Human leukocyte de
Drosophila melanog
Amino acid sequenc
Drosophila melanog
Human antithrombin
Proteinbacterium
Proteinbacterium
Proteinbacterium
Novel human diagn
Streptococcus poly
Cry5F4. Bacillus
Cry5F4. Bacillus
B. thuringiensis c
B. thuringiensis c
Gene #19 associate
Human liver peptid
Human peptide #571
Peptide #598 encod
Protein #559 encod
Human brain expres
Human bone marrow
Peptide #581 encod
Peptide #590 encod
Peptide #567 encod
Human peptide enco
Human ORFX protei
Human ORF628 prote
Escherichia coli s
E. coli cellular p
Mouse RNA-binding
M. tuberculosis an

FT Misc-difference 6 /note= "ATIII.N135A Ile at 390 is substituted by Ala"
 FT FT
 XX XX
 XX WO9958098-A2.
 XX XX
 XX 18-NOV-1999.
 XX XX
 XX 12-MAY-1999; 99WO-US10549.
 XX XX
 XX 12-MAY-1998; 98US-0085197.
 PR 05-MAY-1999; 99US-0085197.
 XX XX
 XX (BOCK/) BOCK S C.
 PA (PICA/) PICARD V.
 PA (ZEND/) ZENDEHROUH P.
 XX XX
 PI Bock SC, Picard V, Zendehtrouh P;
 XX XX
 DR WPI; 2000-116274/10.
 XX XX
 XX New modified human antithrombin III compounds, used for treating e.g.
 PT sepsis, trauma, acute respiratory distress syndrome, restenosis,
 PT thrombosis, thromboembolism or stroke -
 XX XX
 PS Claim 13; Page 57; 75pp; English.
 XX XX
 CC The present sequence is from an antithrombin III (ATIII) variant, 7EVEA
 CC derived from human ATIII.N135A cDNA insert of the pBluebac baculovirus
 CC expression construct and comprises residues 385-393. The variant has
 CC improved resistance to elastase and IGG-activated neutrophils while
 CC retains anti-thrombin and anti-factor Xa activities. It may be
 CC expressed as glycoforms with enhanced heparin affinity which target the
 CC blood vessel wall more efficiently than ATIIIs with normal heparin
 CC affinity. The modified ATIIIs can be used to treat thrombin activation-
 CC related pathological symptoms due to sepsis, trauma, acute
 CC respiratory distress syndrome, restenosis, thrombosis, thromboembolism
 CC and stroke. It can also be used to reduce the risk of reocclusion
 CC and restenosis in percutaneous transluminal coronary angioplasty,
 CC thrombosis associated with surgery, ischaemia/reperfusion injury, and
 CC coagulation abnormalities in cancer or surgical patients.
 XX XX
 SQ Sequence 9 AA;
 Query Match 86.4%; Score 38; DB 21; Length 9;
 Best Local Similarity 88.9%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 STEVEGAGR 9
 |||||
 Db 1 STEVEAAGR 9
 |||||
 RESULT 4
 AAY44473
 ID AAY44473 standard; peptide; 9 AA.
 XX XX
 AC AAY44473;
 XX XX
 DT 27-MAR-2000 (first entry)
 XX XX
 DE Human antithrombin III variant 13.A (residues 385-393).
 XX XX
 KW Human; antithrombin III; ATIII variant 13.A; elastase-resistant;
 KW IGG activated neutrophil resistant; anti-thrombin activity; heparin;
 KW anti-factor Xa activity; blood clotting disorder; sepsis; trauma; stroke;
 KW thrombin activation-related pathological symptom; restenosis; thrombosis;
 KW acute respiratory distress syndrome; thromboembolism; reocclusion.
 XX XX
 OS Homo sapiens.
 OS Synthetic.
 XX XX
 FH Key Location/Qualifiers
 FT Misc-difference 5

FT Misc-difference 6 /note= "ATIII.N135A Val at 389 is substituted by Glu"
 FT FT
 XX XX
 XX WO9958098-A2.
 XX XX
 XX 18-NOV-1999.
 XX XX
 XX 12-MAY-1999; 99WO-US10549.
 XX XX
 XX 12-MAY-1998; 98US-0085197.
 PR 05-MAY-1999; 99US-0085197.
 XX XX
 XX (BOCK/) BOCK S C.
 PA (PICA/) PICARD V.
 PA (ZEND/) ZENDEHROUH P.
 XX XX
 PI Bock SC, Picard V, Zendehtrouh P;
 XX XX
 DR WPI; 2000-116274/10.
 XX XX
 XX New modified human antithrombin III compounds, used for treating e.g.
 PT sepsis, trauma, acute respiratory distress syndrome, restenosis,
 PT thrombosis, thromboembolism or stroke -
 XX XX
 PS Claim 13; Page 57; 75pp; English.
 XX XX
 CC The present sequence is from an antithrombin III (ATIII) variant, 13.A
 CC derived from human ATIII.N135A cDNA insert of the pBluebac baculovirus
 CC expression construct and comprises residues 385-393. The variant has
 CC improved resistance to elastase and IGG-activated neutrophils while
 CC retains anti-thrombin and anti-factor Xa activities. It may be
 CC expressed as glycoforms with enhanced heparin affinity which target the
 CC blood vessel wall more efficiently than ATIIIs with normal heparin
 CC affinity. The modified ATIIIs can be used to treat thrombin activation-
 CC related pathological symptoms due to sepsis, trauma, acute
 CC respiratory distress syndrome, restenosis, thrombosis, thromboembolism
 CC and stroke. It can also be used to reduce the risk of reocclusion
 CC and restenosis in percutaneous transluminal coronary angioplasty,
 CC thrombosis associated with surgery, ischaemia/reperfusion injury, and
 CC coagulation abnormalities in cancer or surgical patients.
 XX XX
 SQ Sequence 9 AA;
 Query Match 86.4%; Score 38; DB 21; Length 9;
 Best Local Similarity 88.9%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 STEVEGAGR 9
 |||||
 Db 1 STAVEGAGR 9
 |||||
 RESULT 5
 AAY68878
 ID AAY68878 standard; Protein; 359 AA.
 XX XX
 AC AAY68878;
 XX XX
 DT 16-MAY-2000 (first entry)
 XX XX
 DE A murine mu-opioid receptor splice variant MOR-1G.
 XX XX
 KW Mu-opioid receptor; MOR-1; splice variant; morphine analgesia;
 KW opioid-mediated ingestive response; opioid activity; analgesic;
 KW gastrointestinal motility; respiration; immune system;
 KW endocrine system; autonomous nervous system; peristalsis regulator;
 KW body weight; neuroendocrine disorder; MOR-1G.
 XX XX
 OS Mus sp.
 XX XX
 FN WO200004046-A2.
 XX XX

XX WO9958098-A2.
XX 18-NOV-1999.
XX
XX 12-MAY-1999; 99WO-US10549.
XX 12-MAY-1998; 98US-0085197.
XX 05-MAY-1999; 99US-0085197.
XX (BOCK/) BOCK S C.
XX (PICA/) PICARD V.
XX (ZEND/) ZENDEHROUH P.
XX Bock SC, Picard V, Zendeherouh P;
XX WPI; 2000-116274/10.
XX New modified human antithrombin III compounds, used for treating e.g. sepsis, trauma, acute respiratory distress syndrome, restenosis, thrombosis, thromboembolism or stroke -
XX
XX Claim 13; Page 57; 75pp; English.
XX The present sequence is from an antithrombin III (ATIII) variant, 13.D derived from human ATIII.N135A cDNA insert of the pBluebac baculovirus expression construct and comprises residues 385-393. The variant has improved resistance to elastase and IGG-activated neutrophils while retains anti-thrombin and anti-factor Xa activities. It may be expressed as glycoforms with enhanced heparin affinity which target the blood vessel wall more efficiently than ATIIIs with normal heparin affinity. The modified ATIIIs can be used to treat thrombin activation-related pathological symptoms due to sepsis, trauma, acute respiratory distress syndrome, restenosis, thrombosis, thromboembolism and stroke. It can also be used to reduce the risk of reocclusion and restenosis in percutaneous transluminal coronary angioplasty, CC thrombosis associated with surgery, ischaemia/reperfusion injury, and coagulation abnormalities in cancer or surgical patients.
XX
XX Sequence 9 AA;
SQ
Query Match 79.5%; Score 35; DB 21; Length 9;
Best Local Similarity 77.8%; Pred. No. 9.3e+05;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 STEVEGAGR 9
DB 1 STALEGAGR 9
RESULT 8
AAU33292
ID AAU33292 standard; Protein; 529 AA.
XX
XX AC AAU33292;
XX
XX 18-DEC-2001 (first entry)
XX
XX Novel human secreted protein #3783.
XX
XX Human; vaccination; gene therapy; nutritional supplement;
KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
XX Homo sapiens.
XX WO200179449-A2.
XX
XX 25-OCT-2001.
XX
XX 16-APR-2001; 2001WO-US08656.
XX
XX 18-APR-2000; 2000US-0552929.
XX

PR 26-JAN-2001; 2001US-0770160.
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT;
XX WPI; 2001-611725/70.
XX
XX Nucleic acids encoding a range of human polypeptides, useful in genetic vaccination, testing and therapy -
XX
XX Claim 20; Page 755-756; 765pp; English.
XX The invention relates to novel human secreted polypeptides. The polypeptides and antibodies to the polypeptides are useful for determining the presence of or predisposition to a disease associated with altered levels of polypeptide. The polypeptides are also useful for identifying agents (agonists and antagonists) that bind to them. Cells expressing the proteins are useful for identifying a therapeutic agent for use in treatment of a pathology related to aberrant expression or physiological interactions of the polypeptide. Vectors comprising the nucleic acids encoding the polypeptides and cells genetically engineered to express them are also useful for producing the proteins. The proteins are useful in genetic vaccination, testing and therapy, and can be used as nutritional supplements. They may be used to increase stem cell proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon and/or nerve tissue growth or regeneration; immune suppression and/or stimulation; as anti-inflammatory agents; and in treatment of leukaemias. AAU29510-AAU3304 represent the amino acid sequences of novel human secreted proteins of the invention.
XX
XX Sequence 529 AA;
SQ
Query Match 79.5%; Score 35; DB 22; Length 529;
Best Local Similarity 87.5%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 TEVEGAGR 9
DB 12 TEVEGGQR 19
RESULT 9
AAU44470
ID AAU44470 standard; peptide; 9 AA.
XX
XX AC AAU44470;
XX
XX 27-MAR-2000 (first entry)
XX
XX Human antithrombin III variant Bb.B (385-393).
XX
XX Human; antithrombin III; ATIII variant Bb.B; elastase-resistant; IGG activated neutrophil resistant; anti-thrombin activity; heparin; anti-factor Xa activity; blood clotting disorder; sepsis; trauma; stroke; thrombin activation-related pathological symptom; restenosis; thrombosis; acute respiratory distress syndrome; thromboembolism; reocclusion.
XX
XX Homo sapiens.
XX Synthetic.
XX
XX Key Location/Qualifiers
FH Misc-difference 3 /note= "ATIII.N135A Ala at 387 is substituted by Glu"
FT
FT Misc-difference 4 /note= "ATIII.N135A Val at 388 is substituted by Gly"
FT
FT Misc-difference 5 /note= "ATIII.N135A Val at 389 is substituted by Glu"
FT
FT Misc-difference 6 /note= "ATIII.N135A Ile at 390 is substituted by Gly"
FT
FT Misc-difference 7 /note= "ATIII.N135A Ala at 391 is substituted by Ser"
XX

PT New system for leukocyte expression profiling, diagnosing a disease, or
PT monitoring (the rate of) progression of a disease, e.g. atherosclerosis
PT or congestive heart failure, comprises diagnostic oligonucleotides
XX
XX Example 5; Page 110; 2038pp; English.

XX The invention relates to a system for detecting gene expression, which
CC comprises one or two isolated DNA molecules that detect expression of a
CC gene, where the gene corresponds to any of 8143 oligonucleotides
CC (ABZ00010-ABZ08152) each having 50 base pairs (bp). The system is useful
CC for leukocyte expression profiling. It is particularly useful for
CC diagnosing a disease, monitoring (rate of) progression of a disease,
CC predicting disease outcome, determining prognosis for a patient,
CC predicting disease complications in an individual or monitoring response
CC to treatment in an individual. The diseases include cardiac allograft
CC rejection, kidney allograft rejection, liver allograft rejection,
CC atherosclerosis, congestive heart failure, systemic lupus erythematosus,
CC rheumatoid arthritis, osteoarthritis or cytomegalovirus infection. The
CC present sequence is that of a human leukocyte expressed protein related
CC to the invention.

XX Sequence 730 AA;
Query Match 77.3%; Score 34; DB 23; Length 730;
Best Local Similarity 87.5%; Pred. No. 3.4e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 STEVEGAG 8
Db 313 STEVRGAG 320
|||||

RESULT 12
ID ABB68486 standard; Protein; 294 AA.
XX
AC ABB68486;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 32250.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI: 2001-656860/75.
DR N-PSDB; ABL12589.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Disclosure; SEQ ID NO 32250; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABBS7737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 294 AA;
Query Match 75.0%; Score 33; DB 22; Length 294;
Best Local Similarity 66.7%; Pred. No. 2e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 STEVEGAG 9
Db 254 TTDVSGAGR 262
:|:|:|

RESULT 13
ID AAB48240 standard; Protein; 514 AA.
XX
AC AAB48240;
XX
DT 02-APR-2001 (first entry)
XX

DE Amino acid sequence of bZIP1 ORF3 protein.

XX Transcription factor; seed storage protein; lectin; oil-body protein;
KW Pv-seed factor-1; ROM1; Vicilin-box binding protein-1; ROM2; 7S-globulin;
KW phaseolin; PHA-L; bean; nuclear protein; promoter; bZIP; ORF;
KW basic leucine zipper.
XX
OS Phaseolus vulgaris.

XX Key Location/Qualifiers
FH Misc-difference 1..514 /note= "Xaa are residues encoded by internal stop codons"
FT FT Misc-difference 432 /note= "encoded by GAC"
XX

PN US6160202-A.

XX 12-DEC-2000.

XX 06-FEB-1997; 97US-0796899.

XX 07-OCT-1994; 94US-0319544.

XX (UYMA-) UNIV MARYLAND BALTIMORE COUNTY.

XX Chern M, Bustos MM;

XX WPI: 2001-079619/09.

XX N-PSDB; AAC84564.

XX Novel transcription factor gene which encodes transcription factor
PT protein that targets promoters of genes encoding seed storage proteins
PT are useful for modulating seed storage protein expression in dicot seed
PT crops -
XX
XX Disclosure; Columns 27-30; 67pp; English.

XX The invention relates to an isolated transcription factor gene which is
CC expressed in a recombinant maturing dicot seed and which encodes a
CC transcription factor protein which targets a promoter of a gene encoding
CC seed storage proteins, lectins or oil-body proteins. The transcription
CC factors isolated are Pv-seed factor-1 (ROM1) and Vicilin-box binding
CC protein-1 (ROM2). These factors bind to 7S-globulin (b-phaseolin) or
CC lectin (PHA-L) promoters. The transcription factor gene is useful for
CC enhancing or reducing expression of seed storage protein, lectin or
CC oil-protein genes in dicot seed crops. The present sequence represents

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 18:21:01 ; Search time 8.22222 Seconds
(without alignments)
46.313 Million cell updates/sec

Title: US-10-014-658-10

Perfect score: 44

Sequence: 1 STEVEGAG 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pdp:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pdp:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pdp:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pdp:*
5: /cgn2_6/ptodata/2/iaa/PTUS_COMB.pdp:*
6: /cgn2_6/ptodata/2/iaa/backfiles.pdp:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36	81.8	359	4 US-09-761-962A-18	Sequence 18, Appl
2	36	81.8	399	4 US-09-761-962A-21	Sequence 21, Appl
3	33	75.0	514	3 US-08-796-895-25	Sequence 25, Appl
4	32	72.7	1167	1 US-08-100-703-2	Sequence 2, Appl
5	32	72.7	1167	1 US-08-176-863-2	Sequence 2, Appl
6	32	72.7	1167	1 US-08-474-038-2	Sequence 2, Appl
7	32	72.7	1167	2 US-08-779-046-2	Sequence 2, Appl
8	32	72.7	1167	2 US-08-881-340-2	Sequence 2, Appl
9	31	70.5	207	2 US-08-870-518-9	Sequence 9, Appl
10	31	70.5	215	4 US-09-291-170A-6	Sequence 6, Appl
11	31	70.5	215	4 US-09-724-884-6	Sequence 6, Appl
12	31	70.5	297	4 US-09-252-991A-28307	Sequence 28307, A
13	31	70.5	405	2 US-08-222-719-2	Sequence 2, Appl
14	31	70.5	405	2 US-08-470-925-2	Sequence 2, Appl
15	31	70.5	405	2 US-08-471-613-2	Sequence 2, Appl
16	31	70.5	405	5 PCT-US93-10443-2	Sequence 2, Appl
17	31	70.5	459	2 US-08-870-518-1	Sequence 1, Appl
18	31	70.5	459	2 US-08-870-518-2	Sequence 2, Appl
19	31	70.5	517	4 US-09-252-991A-31115	Sequence 31115, A
20	30	68.2	112	3 US-08-899-330-3	Sequence 3, Appl
21	30	68.2	112	3 US-08-899-330-4	Sequence 4, Appl
22	30	68.2	124	4 US-09-328-352-5437	Sequence 5437, Ap
23	30	68.2	180	4 US-09-252-991A-23744	Sequence 23744, A
24	30	68.2	209	2 US-08-993-228-6	Sequence 6, Appl
25	30	68.2	223	4 US-09-252-991A-16870	Sequence 16870, A
26	30	68.2	276	4 US-09-252-991A-29800	Sequence 29800, A
27	30	68.2	286	4 US-09-252-991A-20535	Sequence 20535, A

ALIGNMENTS

RESULT 1

US-09-761-962A-18
; Sequence 18, Application US/09761962A
; Patent No. 6500927
; GENERAL INFORMATION:
; APPLICANT: Memorial Sloan-Kettering Cancer Center
; TITLE OF INVENTION: MULTIPLE SPLICE VARIANTS OF THE MU-OPIOID RECEPTOR GENE
; FILE REFERENCE: 830002-2000.2
; CURRENT APPLICATION NUMBER: US/09/761,962A
; CURRENT FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/743,872
; PRIOR APPLICATION NUMBER: PCT/US99/15974
; PRIOR FILING DATE: 1999-07-15
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 18
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-761-962A-18

Query Match 81.8%; Score 36; DB 4; Length 359;

Best Local Similarity 87.5%; Pred. No. 18; Mismatches 0; Indels 0; Gaps 0;

Matches 7; Conservative 1;

Qy 1 STEVEGAG 8

Db 19 STELEGAG 26

RESULT 2

US-09-761-962A-21
; Sequence 21, Application US/09761962A
; Patent No. 6500927
; GENERAL INFORMATION:
; APPLICANT: Memorial Sloan-Kettering Cancer Center
; TITLE OF INVENTION: MULTIPLE SPLICE VARIANTS OF THE MU-OPIOID RECEPTOR GENE
; FILE REFERENCE: 830002-2000.2
; CURRENT APPLICATION NUMBER: US/09/761,962A
; CURRENT FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/743,872
; PRIOR APPLICATION NUMBER: PCT/US99/15974
; PRIOR FILING DATE: 1999-07-15
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
; LENGTH: 399
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-761-962A-21

```

; APPLICATION NUMBER: US 08/100,709
; FILING DATE: 29-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Egolf, Christopher
; REGISTRATION NUMBER: 27633
; REFERENCE/DOCKET NUMBER: 7205-49
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-757-1590
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1167 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-176-865-2

Query Match 72.7%; Score 32; DB 1; Length 1167;
Best Local Similarity 75.0%; Pred. No. 3.8e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 TEVEGAGR 9
DB 97 TELEGLGR 104

RESULT 6
US-08-474-038-2
; Sequence 2, Application US/08474038
; Patent No. 5679343
; GENERAL INFORMATION:
; APPLICANT: Donovan, William P.
; APPLICANT: Tan, Yuping
; APPLICANT: Jany, Christine S.
; APPLICANT: Gonzalez Jr., Jose M.
; TITLE OF INVENTION: BACILLUS THURINGIENSIS CRYET4 AND CRYET5
; TITLE OF INVENTION: TOXIN GENES AND PROTEINS TOXIC TO LEPIDOPTERAN INSECTS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Panitch Schwarze Jacobs & Nadel c/o A.S.
; STREET: 1601 Market Street, 36th Floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,038
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/176,865
; FILING DATE: 30-DEC-1993
; APPLICATION NUMBER: US 08/100,709
; FILING DATE: 29-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Egolf, Christopher
; REGISTRATION NUMBER: 27633
; REFERENCE/DOCKET NUMBER: 7205-49
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-757-1590
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1167 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-474-038-2

QY 2 TEVEGAGR 9
DB 97 TELEGLGR 104

RESULT 7
US-08-779-046-2
; Sequence 2, Application US/08779046
; Patent No. 5854053
; GENERAL INFORMATION:
; APPLICANT: Donovan, William P.
; APPLICANT: Tan, Yuping
; APPLICANT: Jany, Christine S.
; APPLICANT: Gonzalez Jr., Jose M.
; TITLE OF INVENTION: BACILLUS THURINGIENSIS CRYET4 AND CRYET5
; TITLE OF INVENTION: TOXIN GENES AND PROTEINS TOXIC TO LEPIDOPTERAN INSECTS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Panitch Schwarze Jacobs & Nadel c/o A.S.
; STREET: 1601 Market Street, 36th Floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/779,046
; FILING DATE: 06-JAN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/100,709
; FILING DATE: 29-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Egolf, Christopher
; REGISTRATION NUMBER: 27633
; REFERENCE/DOCKET NUMBER: 7205-49
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-757-1590
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1167 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-779-046-2

Query Match 72.7%; Score 32; DB 2; Length 1167;
Best Local Similarity 75.0%; Pred. No. 3.8e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 TEVEGAGR 9
DB 97 TELEGLGR 104

RESULT 8
US-08-881-340-2
; Sequence 2, Application US/08881340
; Patent No. 5942658
; GENERAL INFORMATION:
; APPLICANT: Donovan, William P.
; APPLICANT: Tan, Yuping
; APPLICANT: Jany, Christine S.
; APPLICANT: Gonzalez Jr., Jose M.
; TITLE OF INVENTION: BACILLUS THURINGIENSIS CRYET4 AND CRYET5
; TITLE OF INVENTION: TOXIN GENES AND PROTEINS TOXIC TO LEPIDOPTERAN INSECTS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Panitch Schwarze Jacobs & Nadel c/o A.S.
; STREET: 1601 Market Street, 36th Floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/881,340
; FILING DATE: 06-JAN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/100,709
; FILING DATE: 29-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Egolf, Christopher
; REGISTRATION NUMBER: 27633
; REFERENCE/DOCKET NUMBER: 7205-49
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-757-1590
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1167 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-881-340-2
```



```
; PRIOR APPLICATION NUMBER: 09/291,170
; PRIOR FILING DATE: 1999-04-13
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:
; OTHER INFORMATION: AAA ATPase superfamily sugl AAA domain
US-09-724-884-6

Query Match          70.5%; Score 31; DB 4; Length 215;
Best Local Similarity 75.0%; Pred. No. 1e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 STEVEGAG 8
      |||||
Db      111 STRVEGSG 118

RESULT 12
US-09-252-991A-28307
; Sequence 28307, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 28307
; LENGTH: 297
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28307

Query Match          70.5%; Score 31; DB 4; Length 297;
Best Local Similarity 85.7%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3 EVEGAGR 9
      |||||
Db      270 EVEGVGR 276

RESULT 13
US-08-222-719-2
; Sequence 2, Application US/08222719
; Patent No. 5846711
; GENERAL INFORMATION:
; APPLICANT: David D. Moore
; TITLE OF INVENTION: NUCLEAR HORMONE RECEPTOR-
; TITLE OF INVENTION: INTERACTING POLYPEPTIDES AND
; TITLE OF INVENTION: RELATED MOLECULES AND METHODS
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50Z or 55SX
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,925
; FILING DATE: 06-June-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/222,719
; FILING DATE: 04-April-1994
; CLASSIFICATION: 435
; APPLICATION NUMBER: 07/969,136
; FILING DATE: October 30, 1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul T. Clark
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00786/229003
```

```
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/222,719
; FILING DATE: 04-April-1994
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/969,136
; FILING DATE: 30-October-1992
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul T. Clark
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00786/229001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 405
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-08-222-719-2

Query Match          70.5%; Score 31; DB 2; Length 405;
Best Local Similarity 75.0%; Pred. No. 2e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 STEVEGAG 8
      |||||
Db      255 STRVEGSG 262

RESULT 14
US-08-470-925-2
; Sequence 2, Application US/08470925
; Patent No. 5866886
; GENERAL INFORMATION:
; APPLICANT: David D. Moore
; APPLICANT: Jae Woon Lee
; TITLE OF INVENTION: NUCLEAR HORMONE RECEPTOR-
; TITLE OF INVENTION: INTERACTING POLYPEPTIDES AND
; TITLE OF INVENTION: RELATED MOLECULES AND METHODS
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50Z or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,925
; FILING DATE: 06-June-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/222,719
; FILING DATE: 04-April-1994
; CLASSIFICATION: 435
; APPLICATION NUMBER: 07/969,136
; FILING DATE: October 30, 1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul T. Clark
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00786/229003
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 18:21:27 ; Search time 49.3333 Seconds
(without alignments)
33.929 Million cell updates/sec

Title: US-10-014-658-10

Perfect score: 44

Sequence: 1 STEVEGAGR 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 684280 seqs, 185983659 residues

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36	81.8	359	9	US-09-761-962-18
2	36	81.8	359	15	US-10-283-300-18
3	36	81.8	399	9	US-09-761-962-21
4	36	81.8	399	15	US-10-283-300-21
5	34	77.3	135	10	US-09-738-626-4481
6	33	75.0	314	15	US-10-156-761-13895
7	33	75.0	2174	12	US-10-087-887-87
8	32	72.7	291	15	US-10-156-761-9264
9	31	70.5	42	9	US-09-726-643-144
10	31	70.5	42	14	US-10-042-141-144
11	31	70.5	54	9	US-09-864-761-33858
12	31	70.5	127	9	US-09-815-242-10432
13	31	70.5	178	16	US-10-080-170-187
14	31	70.5	230	10	US-09-988-915-6
15	31	70.5	259	9	US-09-815-242-11976

16	31	70.5	263	15	US-10-156-761-13669
17	31	70.5	267	9	US-09-726-643-143
18	31	70.5	267	14	US-10-042-141-143
19	31	70.5	296	10	US-09-925-300-1260
20	31	70.5	325	11	US-09-934-455-138
21	31	70.5	325	12	US-10-225-068-246
22	31	70.5	429	9	US-09-815-242-14022
23	31	70.5	459	10	US-09-988-315-1
24	31	70.5	459	12	US-10-205-119-111
25	31	70.5	1461	15	US-10-021-955-88
26	30	68.2	52	15	US-10-106-698-6985
27	30	68.2	57	15	US-10-125-358-56
28	30	68.2	63	15	US-10-125-358-56
29	30	68.2	85	15	US-10-060-830-9
30	30	68.2	108	11	US-09-764-921-2719
31	30	68.2	108	15	US-10-205-428-253
32	30	68.2	122	15	US-10-067-974-14
33	30	68.2	213	15	US-10-187-267A-65
34	30	68.2	250	10	US-09-738-626-5662
35	30	68.2	250	11	US-09-746-660A-36
36	30	68.2	250	15	US-10-067-974-10
37	30	68.2	270	12	US-10-214-473-64
38	30	68.2	270	12	US-10-272-490-64
39	30	68.2	275	15	US-10-156-761-8879
40	30	68.2	284	9	US-09-925-301-1267
41	30	68.2	360	12	US-10-029-386-32422
42	30	68.2	384	15	US-10-272-419-6
43	30	68.2	395	15	US-10-156-761-11994
44	30	68.2	428	9	US-09-815-242-11865
45	30	68.2	441	15	US-10-156-761-13669

ALIGNMENTS

RESULT 1

US-09-761-962-18
; Sequence 18, Application US/09761962
; Patent No. US20020077285A1
; GENERAL INFORMATION:
; APPLICANT: Memorial Sloan-Kettering Cancer Center
; TITLE OF INVENTION: Identification and Characterization of Multiple Splice Variants of Mu-
; TITLE OF INVENTION: Opioid Receptor (MOR-1) Gene
; FILE REFERENCE: 830002-2000.1
; CURRENT APPLICATION NUMBER: US/09/761,962
; CURRENT FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 09/743,872
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 18
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-761-962-18

Query Match 81.8%; Score 36; DB 9; Length 359;
Best Local Similarity 87.5%; Pred. No. 55;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 STEVEGAG 8

Db 19 STELEGAG 26

RESULT 2

US-10-283-300-18
; Sequence 18, Application US/10283300
; Publication No. US20030103972A1
; GENERAL INFORMATION:
; APPLICANT: Memorial Sloan-Kettering Cancer Center
; TITLE OF INVENTION: IDENTIFICATION AND CHARACTERIZATION OF MULTIPLE SPLICE VARIANTS

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 13895
; LENGTH: 314
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-13895

Query Match 75.0%; Score 33; DB 15; Length 314;
Best Local Similarity 66.7%; Pred. No. 1.8e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 STEVEGAGR 9
; : : : : :
Db 211 AAEVDGAGR 219

RESULT 7
US-10-087-887-87
; Sequence 87, Application US/10087887
; Publication No. US20030198957A1
; GENERAL INFORMATION:
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Conley, Pamela B.
; APPLICANT: Yang, Ruey-Bing
; APPLICANT: Hart, Matthew
; APPLICANT: Tomlinson, James E.
; APPLICANT: Topper, James N.
; APPLICANT: Shinkets, Richard A.
; APPLICANT: Leach, Martin D.
; APPLICANT: Zerhusen, Bryan D.
; APPLICANT: Komuves, Jaszi
; APPLICANT: Padigaru, Muralidhara
; TITLE OF INVENTION: PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-285
; CURRENT APPLICATION NUMBER: US/10/087,887
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: 60/273,049
; PRIOR FILING DATE: 2001-03-02
; PRIOR APPLICATION NUMBER: 60/279,883
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/277,791
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/281,248
; PRIOR FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: 60/282,864
; PRIOR FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: 60/282,537
; PRIOR FILING DATE: 2001-04-09
; PRIOR APPLICATION NUMBER: 60/282,867
; PRIOR FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 104
; SOFTWARE: CuroSeqList version 0.1
; SEQ ID NO 87
; LENGTH: 2174
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-087-887-87

Query Match 75.0%; Score 33; DB 12; Length 2174;
Best Local Similarity 55.6%; Pred. No. 1.3e+03;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 STEVEGAGR 9
; : : : : :
Db 1349 TTEIGSQ 1357

RESULT 8
US-10-156-761-9264
; Sequence 9264, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9264
; LENGTH: 291
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-9264

Query Match 72.7%; Score 32; DB 15; Length 291;
Best Local Similarity 55.6%; Pred. No. 2.6e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 STEVEGAGR 9
; : : : : :
Db 188 AAEIDGAGR 196

RESULT 9
US-09-726-643-144
; Sequence 144, Application US/09726643
; Patent No. US20020028449A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 26 Human secreted proteins
; FILE REFERENCE: PZ040PI
; CURRENT APPLICATION NUMBER: US/09/726,643
; CURRENT FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: PCT/US00/15187
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: 60/137,725
; PRIOR FILING DATE: 1999-06-07
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 144
; LENGTH: 42
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-726-643-144

Query Match 70.5%; Score 31; DB 9; Length 42;
Best Local Similarity 66.7%; Pred. No. 54;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 STEVEGAGR 9
; : : : : :
Db 30 SSEVDAGR 38

RESULT 10
US-10-042-141-144
; Sequence 144, Application US/10042141
; Publication No. US20020183503A1

; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10432

Query Match 70.5%; Score 31; DB 9; Length 127;
Best Local Similarity 55.6%; Pred. No. 1.7e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 STEVEGAGR 9
: ||| : |||
DB 4 NTEAQSGR 12

RESULT 13

US-10-080-170-187
; Sequence 187, Application US/10080170
; Publication No. US20030129601A1
; GENERAL INFORMATION:
; APPLICANT: COLE, S.T.

; TITLE OF INVENTION: IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
; TITLE OF INVENTION: TREATMENT OF MYCOBACTERIOSES
; FILE REFERENCE: 03495.0218
; CURRENT APPLICATION NUMBER: US/10/080,170

; PRIOR FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 60/270,123
; PRIOR FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 652
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 187
; LENGTH: 178

; TYPE: PRT
; ORGANISM: Mycobacterium leprae
US-10-080-170-187

Query Match 70.5%; Score 31; DB 16; Length 178;
Best Local Similarity 75.0%; Pred. No. 2.4e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TEVEGAGR 9
: ||| |||
DB 136 TSAEGAGR 143

RESULT 14

US-09-988-915-6
; Sequence 6, Application US/09988915
; Patent No. US20020102614A1
; GENERAL INFORMATION:
; APPLICANT: Davis, Roger J.

; TITLE OF INVENTION: USE OF ZPR1 AS A MOLECULAR PROBE FOR SPINAL MUSCULAR ATROPHY
; FILE REFERENCE: 07917-132001
; CURRENT APPLICATION NUMBER: US/09/988,915
; PRIOR FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: US 60/249,745
; PRIOR FILING DATE: 2000-11-17

; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 230
; TYPE: PRT
; ORGANISM: Homo sapiens ZPR1 NH2-terminal domain
US-09-988-915-6

Query Match 70.5%; Score 31; DB 10; Length 230;
Best Local Similarity 55.6%; Pred. No. 3.2e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 STEVEGAGR 9
: ||| : |||
DB 87 NTEIQSAGR 95

RESULT 15

US-09-815-242-11976
; Sequence 11976, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert

; APPLICANT: Chlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11976
; LENGTH: 259
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-11976

Query Match 70.5%; Score 31; DB 9; Length 259;
Best Local Similarity 85.7%; Pred. No. 3.6e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 EVEGAGR 9
: ||| |||
DB 232 EVEGVR 238

Search completed: December 11, 2003, 18:38:04
Job time : 50.3333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 18:19:54 ; Search time 8.3333 Seconds
(without alignments)
103.862 Million cell updates/sec

Title: US-10-014-658-10

Perfect score: 44

Sequence: 1 STEVEGAGR 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_76.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36	81.8	674	2 G70875	probable oxidoredu
2	34	77.3	197	1 A33487	protocatechuate 3,
3	33	75.0	263	2 T35032	probable hydroxyla
4	33	75.0	339	2 T11751	transcription repr
5	32	72.7	291	2 AD2536	DNA polymerase III
6	32	72.7	305	2 AC0828	probable oxidoredu
7	32	72.7	324	2 F71243	probable translati
8	32	72.7	386	2 T12048	ribosomal protein
9	32	72.7	433	2 A29626	apolipoprotein B -
10	32	72.7	610	2 AE2436	penicillin-binding
11	32	72.7	804	2 AF3502	phenylalanine-tRNA
12	32	72.7	837	2 T00618	hypothetical prote
13	31	70.5	110	2 A52448	hypothetical prote
14	31	70.5	127	1 VRECS	preprotein translo
15	31	70.5	127	1 H91241	preprotein translo
16	31	70.5	127	2 E86089	preprotein translo
17	31	70.5	127	2 AC0934	preprotein translo
18	31	70.5	178	2 B87104	conserved hypotet
19	31	70.5	259	2 G83131	conserved hypotet
20	31	70.5	287	2 T36413	probable ABC-type
21	31	70.5	325	2 G86718	unknown protein, 5
22	31	70.5	356	2 T00881	probable PCF2-like
23	31	70.5	368	2 T10557	hypothetical prote
24	31	70.5	373	2 T02976	probable DNA bindi
25	31	70.5	405	2 S64052	26S proteasome reg
26	31	70.5	429	2 AH0630	4-hydroxyphenylace
27	31	70.5	430	2 AF0989	probable membrane
28	31	70.5	454	2 T12539	hypothetical prote
29	31	70.5	679	2 S54299	transketolase (EC

ALIGNMENTS

RESULT 1

G70875
Probable oxidoreductase (EC 1.5.99.-) - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: G70875
R;Cole, S.T.; Broech, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: G70875
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-674 <COL>
A;Cross-references: GB:AL010186; GB:AL123456; NID:G3261493; PIDN:CAA15852.1; PID:G26959
A;Experimental source: strain H37RV
C;Genetics:

A;Gene: fadH
C;Superfamily: Methylotrophus methylotrophus W3A1 trimethylamine dehydrogenase
C;Keywords: 4Fe-4S; metalloprotein; oxidoreductase
F;337,340,344,356/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted
Query Match 81.8%; Score 36; DB 2; Length 674;
Best Local Similarity 87.5%; Pred. No. 19;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 STEVEGAG 8
:|||||
DB 237 ATEVEGAG 244

RESULT 2

A33487
Protocatechuate 3,4-dioxygenase (EC 1.13.11.3) alpha chain - Pseudomonas cepacia
C;Species: Pseudomonas cepacia
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 20-Jan-2003
C;Accession: A33487
R;Zylstra, G.J.; Olsen, R.H.; Ballou, D.P.
J. Bacteriol. 171, 5915-5921, 1989
A;Title: Genetic organization and sequence of the Pseudomonas cepacia genes for the alp
A;Accession: A33487
A;Reference number: A33487; MUID:90036674; PMID:2808303
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-197 <ZYL>
A;Cross-references: GB:M30791; NID:G151429; PIDN:AAA25925.1; PID:G151431
C;Superfamily: protocatechuate 3,4-dioxygenase beta chain
C;Keywords: oxidoreductase

A;Note: this accession replaces an interim accession for a sequence replaced by GenBank
C;Genetics:
A;Gene: PH0208
C;Superfamily: translation initiation factor eIF-2B

Query Match 72.7%; Score 32; DB 2; Length 324;
Best Local Similarity 66.7%; Pred. No. 63;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 STEVEGAGR 9
DB 18 SMEIRGAGR 26
||:|||||

RESULT 8
T12048
ribosomal protein L4 - spoonworm (Urechis caupo)
C;Species: Urechis caupo
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 26-Aug-1999
C;Accession: T12048
R;Rosenthal, E.T.
submitted to the EMBL Data Library, June 1995
A;Reference number: Z17394
A;Accession: T12048
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-386 <ROS>
A;Cross-references: EMBL:U30495; NID:g929960; PID:g929961
C;Superfamily: rat ribosomal protein L4
C;Keywords: ribosome

Query Match 72.7%; Score 32; DB 2; Length 386;
Best Local Similarity 66.7%; Pred. No. 76;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 STEVEGAGR 9
DB 360 ATPVEGAGK 368
||:|||||

RESULT 9
A29626
apolipoprotein B - chicken (fragment)
C;Species: Gallus gallus (chicken)
C;Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 21-Jul-2000
R;Kirchgesner, T.G.; Heinzmann, C.; Svenson, K.L.; Gordon, D.A.; Nicosia, M.; Leberer,
Gene 59, 241-251, 1987
A;Title: Regulation of chicken apolipoprotein B: cloning, tissue distribution, and estr
A;Reference number: A29626; MUID:88137960; PMID:3436530
A;Accession: A29626
A;Molecule type: mRNA
A;Residues: 1-433 <KIR>
A;Cross-references: GB:M18421; NID:g211153; PID:AAA48595.1; PID:g211154
C;Genetics:
A;Gene: apob
C;Keywords: chylomicron; lipid binding; lipoprotein; VLDL

Query Match 72.7%; Score 32; DB 2; Length 433;
Best Local Similarity 75.0%; Pred. No. 86;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 TEVEGAGR 9
DB 253 TDVEGAGR 260
||:|||||

RESULT 10
AE2436
penicillin-binding protein [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002

C;Accession: AE2436
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguch
Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata,
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium An
A;Reference number: AB1807; MUID:21595285; PMID:111759840
A;Accession: AE2436
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-610 <KUR>
A;Cross-references: GB:BA000019; PIDN:BA876744.1; PID:gl7134183; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr5045
C;Superfamily: penicillin-binding protein 3

Query Match 72.7%; Score 32; DB 2; Length 610;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 EVEGAGR 9
DB 228 EVDGAGR 234
||:|||||

RESULT 11
AF3502
phenylalanine-tRNA ligase (EC 6.1.1.20) [imported] - Brucella melitensis (strain 16M)
C;Species: Brucella melitensis
C;Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 03-Jun-2002
C;Accession: AF3502
R;DelVecchio, V.G.; Kaputral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letes
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A;Title: The genome sequence of the facultative intracellular pathogen Brucella meliten
A;Reference number: AD3252; PMID:11756688
A;Accession: AF3502
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-804 <KUR>
A;Cross-references: GB:AB008917; PIDN:AAL53185.1; PID:gl7984058; GSPDB:GN00190
A;Experimental source: strain 16M
C;Genetics:
A;Gene: BMEI2004
A;Map position: 1
C;Superfamily: phenylalanine-tRNA ligase beta chain
C;Keywords: ligase

Query Match 72.7%; Score 32; DB 2; Length 804;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 STEVEGAGR 9
DB 597 TAKVEGAGR 605
||:|||||

RESULT 12
T00618
hypothetical protein T2711.1 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 24-Nov-1999
C;Accession: T00618
R;Federspiel, N.A.; Palm, C.J.; Conway, A.B.; Kurtz, D.B.; Conway, A.R.; Au, M.; Araujo
; Vyotskaia, V.; Yu, G.; Ecker, J.; Theologis, A.; Davis, R.W.
submitted to the EMBL Data Library, September 1998
A;Reference number: Z14193
A;Accession: T00618
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-837 <FED>
A;Cross-references: EMBL:AC004122; NID:g3176693; PID:g3540180; GSPDB:GN00059; ATSP:T27I
A;Experimental source: cultivar Columbia

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 18:19:54 ; Search time 8.3333 Seconds
(without alignments)
103.862 Million cell updates/sec

Title: US-10-014-658-14

Perfect score: 50

Sequence: 1 STQTPNKR 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 9616862 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR 76:**

2: PIR2:**

3: PIR3:**

4: PIR4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	76.0	1054	2 B75384	transcription-repa
2	36	72.0	290	2 AG2216	hypothetical prote
3	36	72.0	341	2 F72501	translation releas
4	36	72.0	351	1 A55362	opsin, pineal glan
5	36	72.0	382	2 S13237	tail fiber protei
6	36	72.0	734	2 T07280	photosystem I P700
7	36	72.0	2142	2 B35098	MHC class III hist
8	35	70.0	130	2 B81951	probable DNA-bind
9	35	70.0	130	2 B81154	hypothetical prote
10	35	70.0	377	2 T24128	hypothetical prote
11	35	70.0	431	2 D89395	protein F3A3.6 li
12	35	70.0	539	2 H84640	probable Rieske ir
13	35	70.0	734	1 A2LVF7	photosystem I P700
14	35	70.0	769	2 B87681	tyrosine kinase D1
15	35	70.0	979	2 T50956	hypothetical prote
16	35	70.0	1233	2 T37045	nitrate reductase
17	35	70.0	1560	2 T00080	hypothetical prote
18	34	68.0	142	2 I51061	MHC class II beta
19	34	68.0	142	2 I51063	MHC class II beta
20	34	68.0	217	2 I51062	MHC class II beta
21	34	68.0	247	2 I51059	MHC class II beta
22	34	68.0	247	2 I51060	MHC class II beta
23	34	68.0	286	2 A55530	megakaryocyte grow
24	34	68.0	326	2 JC4125	thrombopoietin pre
25	34	68.0	336	2 B82064	fructose-1,6-bisph
26	34	68.0	353	2 G32729	thrombopoietin - h
27	34	68.0	353	2 I60105	thrombopoietin pre
28	34	68.0	364	1 PWSFG	H+-transporting tw
29	34	68.0	431	1 B70699	probable pKna prot

30	34	68.0	903	2 T00074	hypothetical prote
31	34	68.0	958	2 T26258	hypothetical prote
32	34	68.0	2655	2 D96595	probable acetyl-Co
33	33	66.0	125	2 G83454	hypothetical prote
34	33	66.0	155	2 B95400	probable transcrip
35	33	66.0	155	2 AF3185	transcription regu
36	33	66.0	204	2 G64754	probable membrane
37	33	66.0	204	2 E90668	probable membrane
38	33	66.0	204	2 H85518	doxC protein - Pse
39	33	66.0	226	2 S27633	Similar to Ubiquit
40	33	66.0	237	2 C86309	Similar to Ubiquit
41	33	66.0	324	2 D84315	aryl-alcohol dehyd
42	33	66.0	330	2 AF3258	hypothetical prote
43	33	66.0	348	2 T28623	hypothetical prote
44	33	66.0	349	2 D72175	G2R protein - vari
45	33	66.0	349	2 D36858	gene G4R protein -

ALIGNMENTS

RESULT 1

B75384
transcription-repair coupling factor - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Nov-2000
C:Accession: B75384
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, B.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; M.
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: B75384
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-1054 <WHI>
A:Cross-references: GB:AE001997; GB:AE00513; MID:G6459292; PIDN:AAF11095.1; PID:G64592
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR1532
A:Map position: 1
A:Superfamily: transcription-repair coupling protein

Query Match 76.0%; Score 38; DB 2; Length 1054;
Best Local Similarity 77.8%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 STQTPNKR 9

DB 689 SIQTFPKGR 697

RESULT 2

AG2216
hypothetical protein all3286 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AG2216
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguch
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, ;
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium An.
A:Reference number: AB1807; MUID:21595285; PMID:111759840
A:Accession: AG2216
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-290 <KUR>
A:Cross-references: GB:BA000019; PIDN:BAE74985.1; PID:gl7132381; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all3286

Bauer, S. C.
Braford-Goldberg, Sarah R.
Caparon, Mairé H.
Easton, Alan M.
Klein, Barbara K.
McKearn, John P.
Olins, Peter O.
Paik, Kuman
Thomas, John W.
TITLE OF INVENTION: Methods Of Ex-Vivo Expansion Of Hematopoietic Cells
Using Multivariant (IL-3) Hematopoiesis Chimera Proteins
NUMBER OF SEQUENCES: 197
CORRESPONDENCE ADDRESS:
ADDRESSEE: S. Christopher Bauer, Pharmacia Corporation
Corporate Patent Dept., Mail Zone O4E
STREET: 800 N. Lindbergh
CITY: St. Louis
STATE: Missouri
COUNTRY: USA
ZIP: 63167
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/083,446
FILING DATE: 26-Feb-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/762,227
FILING DATE: 09-DEC-1996
APPLICATION NUMBER: US 08/192,325
FILING DATE: 14-FEB-1994
APPLICATION NUMBER: US 08/446,872
FILING DATE: 06-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: S. Christopher Bauer
REGISTRATION NUMBER: 42,305
REFERENCE/DOCKET NUMBER: C-2790/6
TELEPHONE: (636)737-6257
TELEFAX: (636)737-5452
INFORMATION FOR SEQ ID NO: 166:
SEQUENCE CHARACTERISTICS:
LENGTH: 286 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 166:
US-10-083-446-166

Query Match 68.0%; Score 34; DB 12; Length 286;
Best Local Similarity 75.0%; Pred. No. 5.2e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TOTPPNGR 9

Db 243 TOLPPQGR 250

RESULT 12

US-10-083-446-167
; Sequence 167, Application US/10083446
; Publication No. US20030185790A1
; GENERAL INFORMATION:
; APPLICANT: Abrams, Mark A.
; Bauer, S. C.
; Braford-Goldberg, Sarah R.
; Caparon, Mairé H.
; Easton, Alan M.
; Klein, Barbara K.

McKearn, John P.
Olins, Peter O.
Paik, Kuman
Thomas, John W.
TITLE OF INVENTION: Methods Of Ex-Vivo Expansion Of Hematopoietic Cells
Using Multivariant (IL-3) Hematopoiesis Chimera Proteins
NUMBER OF SEQUENCES: 197
CORRESPONDENCE ADDRESS:
ADDRESSEE: S. Christopher Bauer, Pharmacia Corporation
Corporate Patent Dept., Mail Zone O4E
STREET: 800 N. Lindbergh
CITY: St. Louis
STATE: Missouri
COUNTRY: USA
ZIP: 63167
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/083,446
FILING DATE: 26-Feb-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/762,227
FILING DATE: 09-DEC-1996
APPLICATION NUMBER: US 08/192,325
FILING DATE: 14-FEB-1994
APPLICATION NUMBER: US 08/446,872
FILING DATE: 06-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: S. Christopher Bauer
REGISTRATION NUMBER: 42,305
REFERENCE/DOCKET NUMBER: C-2790/6
TELEPHONE: (636)737-6257
TELEFAX: (636)737-5452
INFORMATION FOR SEQ ID NO: 167:
SEQUENCE CHARACTERISTICS:
LENGTH: 286 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 167:
US-10-083-446-167

Query Match 68.0%; Score 34; DB 12; Length 286;
Best Local Similarity 75.0%; Pred. No. 5.2e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TOTPPNGR 9

Db 112 TOLPPQGR 119

RESULT 13

US-10-083-446-168
; Sequence 168, Application US/10083446
; Publication No. US20030185790A1
; GENERAL INFORMATION:
; APPLICANT: Abrams, Mark A.
; Bauer, S. C.
; Braford-Goldberg, Sarah R.
; Caparon, Mairé H.
; Easton, Alan M.
; Klein, Barbara K.
; McKearn, John P.
; Olins, Peter O.
; Paik, Kuman
; Thomas, John W.

TITLE OF INVENTION: Methods Of Ex-Vivo Expansion Of Hematopoietic Cells

Query Match 72.0%; Score 36; DB 11; Length 978;
Best Local Similarity 77.8%; Pred. No. 8.1e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 STQTPNGR 9
|:|:|:|:|
DB 617 STQTPPGGR 625

RESULT 6
US-10-156-761-7994
; Sequence 7994, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 7994
; LENGTH: 205
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-7994

Query Match 70.0%; Score 35; DB 15; Length 205;
Best Local Similarity 75.0%; Pred. No. 2.6e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 TQTPNGR 9
|:|:|:|:|
DB 194 TETPANGR 201

RESULT 7
US-10-156-761-14230
; Sequence 14230, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 14230
; LENGTH: 527
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-14230

Query Match 70.0%; Score 35; DB 15; Length 527;

Best Local Similarity 66.7%; Pred. No. 6.5e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 STQTPNGR 9
|:|:|:|:|
DB 3 SVQSPGGR 11

RESULT 8
US-10-083-446-164
; Sequence 164, Application US/10083446
; Publication No. US20030185790A1
; GENERAL INFORMATION:
; APPLICANT: ABRAMS, MARK A.
; APPLICANT: BAUER, S. C.
; APPLICANT: BRAFORD-GOLDBERG, SARAH R.
; APPLICANT: CABERON, MAIRE H.
; APPLICANT: EASTON, ALAN M.
; APPLICANT: KLEIN, BARBARA K.
; APPLICANT: MCKEARN, JOHN P.
; APPLICANT: OLINS, PETER O.
; APPLICANT: PAIK, KUMMAN
; APPLICANT: THOMAS, JOHN W.
; TITLE OF INVENTION: METHODS OF EX-VIVO EXPANSION OF HEMATOPOIETIC CELLS
; USING MULTIVARIANT (IL-3) HEMATOPOIESIS CHIMERA PROTEINS
; NUMBER OF SEQUENCES: 197
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: S. Christopher Bauer, Pharmacia Corporation
; Corporate Patent Dept., Mail Zone 04E
; STREET: 800 N. Lindbergh
; CITY: St. Louis
; STATE: Missouri
; COUNTRY: USA
; ZIP: 63167
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/083,446
; FILING DATE: 26-Feb-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/762,227
; FILING DATE: 09-DEC-1996
; APPLICATION NUMBER: US 08/192,325
; FILING DATE: 14-FEB-1994
; APPLICATION NUMBER: US 08/446,872
; FILING DATE: 06-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: S. Christopher Bauer
; REGISTRATION NUMBER: 42,305
; REFERENCE/DOCKET NUMBER: C-2790/5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (636)737-6257
; TELEFAX: (636)737-5452
; INFORMATION FOR SEQ ID NO: 164:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 155 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 164:
US-10-083-446-164

Query Match 68.0%; Score 34; DB 12; Length 155;
Best Local Similarity 75.0%; Pred. No. 2.9e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TQTPNGR 9
|:|:|:|:|

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 18:21:27 ; Search time 49.3333 Seconds
(without alignments)
33.929 Million cell updates/sec

Title: US-10-014-658-14

Perfect score: 50

Sequence: 1 STQTPPNGR 9

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 684280 seqs, 185983659 residues

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:

1: /cgn2_6/ptodata/1/pubaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubaa/US09C_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	76.0	418	15	US-10-156-761-12646
2	37	74.0	478	15	US-10-156-761-14899
3	36	72.0	322	11	US-09-746-650A-64
4	36	72.0	918	10	US-09-738-626-5661
5	36	72.0	718	11	US-09-938-901-8
6	35	70.0	205	15	US-10-156-761-7994
7	35	70.0	527	15	US-10-156-761-14230
8	34	68.0	155	12	US-10-083-446-164
9	34	68.0	174	12	US-10-072-571-56
10	34	68.0	286	12	US-10-083-446-165
11	34	68.0	286	12	US-10-083-446-166
12	34	68.0	286	12	US-10-083-446-167
13	34	68.0	290	12	US-10-083-446-168
14	34	68.0	332	12	US-10-400-377-7
15	34	68.0	332	12	US-10-400-708-7

16	34	68.0	332	12	US-10-298-148-7
17	34	68.0	353	11	US-09-872-702-13
18	34	68.0	353	11	US-09-774-381-52
19	34	68.0	353	12	US-10-072-571-55
20	34	68.0	353	15	US-10-205-823-409
21	34	68.0	1461	12	US-10-094-749-2733
22	33	66.0	47	9	US-09-864-761-36858
23	33	66.0	90	9	US-09-864-761-37018
24	33	66.0	204	9	US-09-815-242-10054
25	33	66.0	255	11	US-09-784-891-4230
26	33	66.0	349	9	US-09-826-212-13
27	33	66.0	349	10	US-09-935-727-15
28	33	66.0	349	15	US-10-186-643-13
29	33	66.0	400	15	US-10-156-761-11416
30	33	66.0	420	12	US-10-236-055A-2
31	33	66.0	550	12	US-10-303-664A-12
32	33	66.0	743	12	US-10-309-422-34
33	33	66.0	746	10	US-09-738-973-185
34	33	66.0	746	10	US-09-854-133-185
35	33	66.0	746	15	US-10-144-649A-185
36	33	66.0	792	12	US-10-309-422-38
37	33	66.0	827	12	US-10-309-422-42
38	33	66.0	877	12	US-10-309-422-32
39	33	66.0	907	9	US-09-954-043-4
40	33	66.0	907	15	US-10-200-154-4
41	33	66.0	908	12	US-10-309-422-10
42	33	66.0	909	12	US-10-309-422-22
43	33	66.0	926	12	US-10-309-422-36
44	33	66.0	940	15	US-10-102-806-678
45	33	66.0	957	12	US-10-309-422-14

ALIGNMENTS

RESULT 1

US-10-156-761-12646
; Sequence 12646, Application US/10156761
; Publication No. US2003019018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 12646
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-12646

Query Match 76.0%; Score 38; DB 15; Length 418;
Best Local Similarity 77.8%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Cy 1 STQTPPNGR 9
Db 166 STQPPGGR 174

RESULT 2
US-10-156-761-14899

ATTORNEY/AGENT INFORMATION:
NAME: Bennett, Dennis A.
REGISTRATION NUMBER: 34,547
REFERENCE/DOCKET NUMBER: C-2789/S
TELECOMMUNICATION INFORMATION:
TELEPHONE: (708)470-6501
TELEFAX: (708)470-6881
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 174 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-761-907-56

Query Match 68.0%; Score 34; DB 4; Length 174;
Best Local Similarity 75.0%; Pred. No. 98;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TOTPPNGR 9
Db 131 TQLPPQGR 138

RESULT 15
PCT-US95-03776-27
Sequence 27, Application PC/TUS9503776
GENERAL INFORMATION:
APPLICANT: AMGEN INC.
TITLE OF INVENTION: Compositions and Methods for Stimulating
TITLE OF INVENTION: Megakaryocyte Growth and Differentiation
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amgen Inc.
STREET: 1840 Dehavilland Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: USA
ZIP: 91320-1789
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/03776
FILING DATE:
CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:
NAME: Cook, Robert R.
REFERENCE/DOCKET NUMBER: A-290-C
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 174 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-03776-27

Query Match 68.0%; Score 34; DB 5; Length 174;
Best Local Similarity 75.0%; Pred. No. 98;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TOTPPNGR 9
Db 110 TQLPPQGR 117

Search completed: December 11, 2003, 18:30:42
Job time : 9.22222 secs

;; TITLE OF INVENTION: Mutant Polypeptides with CSF's for Multi-lineage
;; TITLE OF INVENTION: Hematopoietic Cell Production
;; NUMBER OF SEQUENCES: 58
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Dennis A. Bennett, G.D. Searle & Co.,
;; ADDRESSEE: Corporate Patent Dept.
;; STREET: P. O. Box 5110
;; CITY: Chicago
;; STATE: Illinois
;; COUNTRY: USA
;; ZIP: 60680
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent In Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/471,045
;; FILING DATE: 06-JUN-1995
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US95/01184
;; FILING DATE: 02-FEB-1995
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/193,373
;; FILING DATE: 04-FEB-1994
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Bennett, Dennis A.
;; REGISTRATION NUMBER: 34,547
;; REFERENCE/DOCKET NUMBER: C-2789/3
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (708)470-6501
;; TELEFAX: (708)470-6881
;; INFORMATION FOR SEQ ID NO: 56:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 174 amino acids
;; TYPE: amino acid
;; STRANDEDNESS:
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-471-045-56

Query Match 68.0%; Score 34; DB 3; Length 174;
Best Local Similarity 75.0%; Pred. No. 98;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TOTPPNGR 9
Db 131 TQLPQGR 138

RESULT 11
US-08-469-712A-56
; Sequence 56, Application US/08469712A
; Patent No. 6093395
; GENERAL INFORMATION:
; APPLICANT: Abrams, Mark A.
; APPLICANT: Bauer, S. C.
; APPLICANT: Braford-Goldberg, Sarah R.
; APPLICANT: Caparon, Mair H.
; APPLICANT: Easton, Alan M.
; APPLICANT: Klein, Barbara K.
; APPLICANT: McKearn, John P.
; APPLICANT: Olin, Peter O.
; APPLICANT: Paik, Kuman
; APPLICANT: Thomas, John W.
; TITLE OF INVENTION: Co-administration of Interleukin-3
; TITLE OF INVENTION: Mutant Polypeptides with CSF's for Multi-lineage
; TITLE OF INVENTION: Hematopoietic Cell Production
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dennis A. Bennett, G.D. Searle & Co.,
; ADDRESSEE: Corporate Patent Dept.

;; STREET: P. O. Box 5110
;; CITY: Chicago
;; STATE: Illinois
;; COUNTRY: USA
;; ZIP: 60680
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent In Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/469,712A
;; FILING DATE: 06-JUN-1995
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US95/01184
;; FILING DATE: 02-FEB-1995
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/193,373
;; FILING DATE: 04-FEB-1994
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Bennett, Dennis A.
;; REGISTRATION NUMBER: 34,547
;; REFERENCE/DOCKET NUMBER: C-2789/2
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (708)470-6501
;; TELEFAX: (708)470-6881
;; INFORMATION FOR SEQ ID NO: 56:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 174 amino acids
;; TYPE: amino acid
;; STRANDEDNESS:
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-469-712A-56

Query Match 68.0%; Score 34; DB 3; Length 174;
Best Local Similarity 75.0%; Pred. No. 98;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TOTPPNGR 9
Db 131 TQLPQGR 138

RESULT 12
US-08-446-871-56
; Sequence 56, Application US/08446871
; Patent No. 6361976
; GENERAL INFORMATION:
; APPLICANT: Abrams, Mark A.
; APPLICANT: Bauer, S. C.
; APPLICANT: Braford-Goldberg, Sarah R.
; APPLICANT: Caparon, Mair H.
; APPLICANT: Easton, Alan M.
; APPLICANT: Klein, Barbara K.
; APPLICANT: McKearn, John P.
; APPLICANT: Olin, Peter O.
; APPLICANT: Paik, Kuman
; APPLICANT: Thomas, John W.
; TITLE OF INVENTION: Co-administration of Interleukin-3
; TITLE OF INVENTION: Mutant Polypeptides with CSF's for Multi-lineage
; TITLE OF INVENTION: Hematopoietic Cell Production
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dennis A. Bennett, G.D. Searle & Co.,
; ADDRESSEE: Corporate Patent Dept.
; STREET: P. O. Box 5110
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60680
; COMPUTER READABLE FORM:

;/ CORRESPONDENCE ADDRESS:
;/ ADDRESSEE: Dennis A. Bennett, G.D. Searle & Co.,
;/ ADDRESSEE: Corporate Patent Dept.,
;/ STREET: P. O. Box 5110
;/ CITY: Chicago
;/ STATE: Illinois
;/ COUNTRY: USA
;/ ZIP: 60680
;/ COMPUTER READABLE FORM:
;/ MEDIUM TYPE: Floppy disk
;/ OPERATING SYSTEM: PC-DOS/MS-DOS
;/ SOFTWARE: Patent In Release #1.0, Version #1.25
;/ CURRENT APPLICATION DATA:
;/ APPLICATION NUMBER: US/08/468,609A
;/ FILING DATE: 06-JUN-1995
;/ CLASSIFICATION:
;/ PRIOR APPLICATION DATA:
;/ APPLICATION NUMBER: US 08/192,325
;/ FILING DATE: 14-FEB-1994
;/ ATTORNEY/AGENT INFORMATION:
;/ NAME: Bennett, Dennis A.
;/ REGISTRATION NUMBER: 34,547
;/ REFERENCE/DOCKET NUMBER: C-2790/3
;/ TELECOMMUNICATION INFORMATION:
;/ TELEPHONE: (314)737-6972
;/ TELEFAX: (314)737-6972
;/ INFORMATION FOR SEQ ID NO: 164:
;/ SEQUENCE CHARACTERISTICS:
;/ LENGTH: 155 amino acids
;/ TYPE: amino acid
;/ STRANDEDNESS:
;/ TOPOLOGY: linear
;/ MOLECULE TYPE: protein
;/ US-08-468-609A-164

Query Match 68.0%; Score 34; DB 3; Length 155;
Best Local Similarity 75.0%; Pred. No. 87;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TOTPPNGR 9
DB 112 TQLPPQGR 119

RESULT 6
US-08-446-872A-164
; Sequence 164, Application US/08446872A
; Patent No. 6361977
; GENERAL INFORMATION:
; APPLICANT: Abrams, Mark A.
; APPLICANT: Bauer, S. C.
; APPLICANT: Braford-Goldberg, Sarah R.
; APPLICANT: Caparon, Mair H.
; APPLICANT: Easton, Alan M.
; APPLICANT: Klein, Barbara K.
; APPLICANT: McKearn, John P.
; APPLICANT: Ollins, Peter O.
; APPLICANT: Paik, Kuman
; APPLICANT: Thomas, John W.
; TITLE OF INVENTION: Multivariant IL-3 Hematopoiesis
; NUMBER OF SEQUENCES: 197
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dennis A. Bennett, G.D. Searle & Co.,
; ADDRESSEE: Corporate Patent Dept.,
; STREET: P. O. Box 5110
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60680
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

;/ COMPUTER: IBM PC compatible
;/ OPERATING SYSTEM: PC-DOS/MS-DOS
;/ SOFTWARE: Patent In Release #1.0, Version #1.25
;/ CURRENT APPLICATION DATA:
;/ APPLICATION NUMBER: US/08/446,872A
;/ FILING DATE: 06-JUN-1995
;/ CLASSIFICATION: 424
;/ PRIOR APPLICATION DATA:
;/ APPLICATION NUMBER: US 08/192,325
;/ FILING DATE: 14-FEB-1994
;/ ATTORNEY/AGENT INFORMATION:
;/ NAME: Bennett, Dennis A.
;/ REGISTRATION NUMBER: 34,547
;/ REFERENCE/DOCKET NUMBER: C-2790/1
;/ TELECOMMUNICATION INFORMATION:
;/ TELEPHONE: (314)737-6986
;/ TELEFAX: (314)737-6972
;/ INFORMATION FOR SEQ ID NO: 164:
;/ SEQUENCE CHARACTERISTICS:
;/ LENGTH: 155 amino acids
;/ TYPE: amino acid
;/ STRANDEDNESS:
;/ TOPOLOGY: linear
;/ MOLECULE TYPE: protein
;/ US-08-446-872A-164
;/ Query Match 68.0%; Score 34; DB 4; Length 155;
;/ Best Local Similarity 75.0%; Pred. No. 87;
;/ Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
;/ QY 2 TOTPPNGR 9
;/ DB 112 TQLPPQGR 119
;/ RESULT 7
;/ US-08-762-227A-164
;/ Sequence 164, Application US/08762227A
;/ Patent No. 6436387
;/ GENERAL INFORMATION:
;/ APPLICANT: Abrams, Mark A.
;/ APPLICANT: Bauer, S. C.
;/ APPLICANT: Braford-Goldberg, Sarah R.
;/ APPLICANT: Caparon, Mair H.
;/ APPLICANT: Easton, Alan M.
;/ APPLICANT: Klein, Barbara K.
;/ APPLICANT: McKearn, John P.
;/ APPLICANT: Ollins, Peter O.
;/ APPLICANT: Paik, Kuman
;/ APPLICANT: Thomas, John W.
;/ TITLE OF INVENTION: Multivariant IL-3 Hematopoiesis
;/ NUMBER OF SEQUENCES: 197
;/ CORRESPONDENCE ADDRESS:
;/ ADDRESSEE: Dennis A. Bennett, G.D. Searle & Co.,
;/ ADDRESSEE: Corporate Patent Dept.,
;/ STREET: P. O. Box 5110
;/ CITY: Chicago
;/ STATE: Illinois
;/ COUNTRY: USA
;/ ZIP: 60680
;/ COMPUTER READABLE FORM:
;/ MEDIUM TYPE: Floppy disk
;/ COMPUTER: IBM PC compatible
;/ OPERATING SYSTEM: PC-DOS/MS-DOS
;/ SOFTWARE: Patent In Release #1.0, Version #1.25
;/ CURRENT APPLICATION DATA:
;/ APPLICATION NUMBER: US/08/762,227A
;/ FILING DATE: 09-Dec-1996
;/ CLASSIFICATION: <Unknown>
;/ PRIOR APPLICATION DATA:
;/ APPLICATION NUMBER: US 08/192,325
;/ FILING DATE: 14-FEB-1994

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 18:21:01 ; Search time 8.22222 Seconds
(without alignments)
46.313 Million cell updates/sec

Title: US-10-014-658-14
Perfect score: 50
Sequence: 1 STQTPNGR 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	70.0	1118	2	US-08-724-354D-2
2	35	70.0	1118	3	US-09-270-984A-2
3	34	68.0	153	3	US-08-875-533-65
4	34	68.0	153	3	US-08-469-318-164
5	34	68.0	155	3	US-08-468-609A-164
6	34	68.0	155	3	US-08-446-872A-164
7	34	68.0	155	4	US-08-762-227A-164
8	34	68.0	155	5	PCT-US95-01185-164
9	34	68.0	165	1	US-08-413-803-29
10	34	68.0	174	3	US-08-471-045-56
11	34	68.0	174	3	US-08-469-712A-56
12	34	68.0	174	4	US-08-446-871-56
13	34	68.0	174	4	US-08-468-910-56
14	34	68.0	174	4	US-08-761-907-56
15	34	68.0	174	5	PCT-US95-03776-27
16	34	68.0	195	1	US-08-388-779A-4
17	34	68.0	195	1	US-08-591-070A-4
18	34	68.0	195	2	US-08-927-855-4
19	34	68.0	265	5	PCT-US95-03776-29
20	34	68.0	265	5	PCT-US95-03776-29
21	34	68.0	285	3	US-08-875-533-69
22	34	68.0	285	3	US-08-875-533-70
23	34	68.0	286	1	US-08-321-488A-27
24	34	68.0	286	3	US-08-469-318-165
25	34	68.0	286	3	US-08-469-318-166
26	34	68.0	286	3	US-08-469-318-167
27	34	68.0	286	3	US-08-468-609A-165

28	34	68.0	286	3	US-08-468-609A-166	Sequence 166, App
29	34	68.0	286	3	US-08-468-609A-167	Sequence 167, App
30	34	68.0	286	4	US-08-446-872A-165	Sequence 165, App
31	34	68.0	286	4	US-08-446-872A-166	Sequence 166, App
32	34	68.0	286	4	US-08-446-872A-167	Sequence 167, App
33	34	68.0	286	4	US-08-762-227A-165	Sequence 165, App
34	34	68.0	286	4	US-08-762-227A-166	Sequence 166, App
35	34	68.0	286	4	US-08-762-227A-167	Sequence 167, App
36	34	68.0	286	5	PCT-US95-01185-165	Sequence 165, App
37	34	68.0	286	5	PCT-US95-01185-166	Sequence 166, App
38	34	68.0	286	5	PCT-US95-01185-167	Sequence 167, App
39	34	68.0	288	3	US-08-875-533-71	Sequence 71, App1
40	34	68.0	288	3	US-08-875-533-73	Sequence 73, App1
41	34	68.0	290	3	US-08-469-318-168	Sequence 168, App
42	34	68.0	290	3	US-08-468-609A-168	Sequence 168, App
43	34	68.0	290	4	US-08-446-872A-168	Sequence 168, App
44	34	68.0	290	4	US-08-762-227A-168	Sequence 168, App
45	34	68.0	290	5	PCT-US95-01185-168	Sequence 168, App

ALIGNMENTS

RESULT 1
US-08-724-354D-2
; Sequence 2, Application US/08724354D
; Patent No. 5994119
; GENERAL INFORMATION:
; APPLICANT: Dietz, Harry C.
; TITLE OF INVENTION: MAMMALIAN REGULATOR OF
; TITLE OF INVENTION: NONSENSE-MEDIATED RNA DECAY
; NUMBER OF INVENTIONS: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,354D
; FILING DATE: 01-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/016,482
; FILING DATE: 29-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07265/090001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-678-5070
; TELEFAX: 619-678-5099
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1118 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-724-354D-2

Query Match 70.0%; Score 35; DB 2; Length 1118;
Best Local Similarity 85.7%; Pred. No. 4.6e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 2 TOTPPNG 8
Db 44 TOTPPG 50

PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 14-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 70.0%; Score 35; DB 21; Length 345;

Best Local Similarity 85.7%; Pred. No. 5e+02;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 STCTPPN 7
 Db 2 STKTPPN 8

RESULT 15

AAG37538
 ID AAG37538 standard; Protein; 345 AA.

XX AC AAG37538;

XX DT 18-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 46171.

XX KW Protein identification; signal transduction pathway; metabolic pathway;
 XX hybridisation assay; genetic mapping; gene expression control; promoter;
 XX termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.
 PR 05-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 01-APR-1999; 99US-0126785.
 PR 06-APR-1999; 99US-0127462.
 PR 08-APR-1999; 99US-0128234.
 PR 15-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 23-APR-1999; 99US-0130891.
 PR 28-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 07-MAY-1999; 99US-0132487.
 PR 11-MAY-1999; 99US-0132863.
 PR 14-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140895.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.

CC engineered to express them are also useful for producing the proteins.
 CC The proteins are useful in genetic vaccination, testing and
 CC therapy, and can be used as nutritional supplements. They may be used to
 CC increase stem cell proliferation; to regulate haematopoiesis; and in
 CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
 CC immune suppression and/or stimulation; as anti-inflammatory agents; and
 CC in treatment of leukaemias. AAU29510-AAU3304 represent the amino acid
 CC sequences of novel human secreted proteins of the invention.

XX Sequence 2153 AA;

Query Match 72.0%; Score 36; DB 22; Length 2153;
 Best Local Similarity 66.7%; Pred. No. 2.1e+03;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 STQTPNGR 9

Db 2059 SRTPTTGR 2067

RESULT 12

AAW36504

ID AAW36504 standard; protein; 57 AA.

XX AC AAW36504;

XX 14-JUL-1998 (first entry)

XX Human RENT1 protein fragment #1.

DE RENT1; nonsense-mediated RNA decay; NMRD; mutation; diagnosis; therapy;

KW regulator of nonsense transcripts; Marfan Syndrome; aging; cancer.

XX Homo sapiens.

XX WO9740855-A1.

XX 06-NOV-1997.

XX 01-OCT-1996; 96WO-US15769.

XX 29-APR-1996; 96US-0016482.

XX (UYJO) UNIV JOHNS HOPKINS SCHOOL MEDICINE.

XX Dietz HC;

XX WPI; 1997-549494/50.

XX New isolated regulators of non-sense-mediated RNA decay - used to

XX develop products for the study, diagnosis and therapy of disorders

XX such as Marfan Syndrome, accelerated ageing and cancers

XX Claim 1; Fig 1A; 79pp; English.

XX This protein sequence is a fragment of the human RENT1 protein which

XX regulates nonsense-mediated RNA decay (NMRD). This fragment of RENT1

XX shows homology to the yeast heat shock protein, Sislp and the Brassica

XX napus RNA binding protein GRP10. The RENT1 (regulator of nonsense

XX transcripts) protein and other products can be used in the study,

XX diagnosis and therapy of disorders involving NMRD such as Marfan

XX Syndrome, accelerated aging or various cancers.

XX Sequence 57 AA;

Query Match 70.0%; Score 35; DB 18; Length 57;

Best Local Similarity 85.7%; Pred. No. 86;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TOTPPNG 8

Db 44 TOTPPGG 50

RESULT 13

AAAY04837

ID AAAY04837 standard; Protein; 203 AA.

XX AC AAAY04837;

XX 06-JUL-1999 (first entry)

XX Mycobacterium species protein sequence 15F.

DE Secreted protein; Mycobacterium; primer; PCR; amplification; probe;

KW hybridisation; detection; vaccine; immunisation; infection.

XX Mycobacterium sp.

XX WO9909186-A2.

XX 25-FEB-1999.

XX 14-AUG-1998; 98WO-PRO1813.

XX 11-SEP-1997; 97FR-0011325.

XX 14-AUG-1997; 97FR-0010404.

XX (INSP) INST PASTEUR.

XX Gicquel B, Lim EM, Pellicic V, Portnoi D, Goguet de la Salmoniere Y;

XX Guigueno A;

XX WPI; 1999-181045/15.

XX N-PSDB; AAX34089.

XX Mycobacterial DNA vectors containing reporter constructs - for

XX identifying coding or promoter sequences involved in

XX infection-associated protein expression

XX Claim 32; Fig 15F; 309pp; French.

XX Sequences AAAY04742-Y05000 and AAAY07201-Y07204 represent secreted

XX proteins from various Mycobacterium species microorganisms. The

XX encoding nucleotide sequences can be used as primers and probes for

XX methods for detecting and identifying mycobacteria, especially belonging

XX to the M. tuberculosis complex. The encoded proteins can be used in

XX vaccines for immunisation against a bacterial or viral infection.

XX Query Match 70.0%; Score 35; DB 20; Length 203;

Best Local Similarity 66.7%; Pred. No. 3e+02;

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 STQTPNGR 9

Db 76 SARTPPGGR 84

RESULT 14

AAAG07022

ID AAAG07022 standard; Protein; 345 AA.

XX AC AAAG07022;

XX 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 4014.

DE Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; Genetic mapping; gene expression control; promoter;

XX termination sequence.

XX Arabidopsis thaliana.

XX Thermus thermophilus; DNA repair enzyme; enzyme; MutY; RecJ; RecF;
 KW TRCF; biochemistry; molecular biology; research.
 XX Thermus thermophilus.
 OS JP2002247985-A.
 PN 03-SEP-2002.
 XX 23-FEB-2001; 2001JP-0047762.
 PF 23-FEB-2001; 2001JP-0047762.
 XX (RIKA) RIKAGAKU KENKYUSHO.
 PA WPI; 2003-078924/08.
 XX DR N-PSDB; AB222146.
 DR A DNA repair enzyme gene, a protein, a recombinant vector, a
 PT transformant, preparation of DNA repair enzyme, repairing the error
 PT sequence of a DNA, and prevention of error synthesis of a DNA sequence
 PT
 XX Claim 1; Page 26-28; 41pp; Japanese.
 PS AB222143 to AB222146 encode the Thermus thermophilus DNA repair enzymes
 CC MutY, RecJ, RecF, and TRCF given in ABP56413 to ABP56416. The enzymes
 CC can be used as research reagents for biochemistry and molecular biology.
 CC
 XX Sequence 978 AA;
 SQ
 Query Match 72.0%; Score 36; DB 24; Length 978;
 Best Local Similarity 77.8%; Pred. No. 9.5e+02;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 STQTPPPNGR 9
 DB 617 SIQTPPPGR 625
 RESULT 10
 ID ABB59227 standard; Protein; 1480 AA.
 AC ABB59227;
 XX 26-MAR-2002 (first entry)
 DT Drosophila melanogaster polypeptide SEQ ID NO 4473.
 DE Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX Drosophila melanogaster.
 OS WO200171042-A2.
 XX 27-SEP-2001.
 PD 23-MAR-2001; 2001WO-US09231.
 XX 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX (PEKE) PE CORP NY.
 PA Venter JC, Adams M, Li FWD, Myers EW;
 XX WPI; 2001-656860/75.
 DR N-PSDB; ABL03330.
 XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX Disclosure; SEQ ID NO 4473; 21pp + Sequence Listing; English.
 PS The invention relates to an isolated nucleic acid detection reagent
 XX capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
 CC sequences (AB101840-AB116175) and the encoded proteins
 CC (AB57737-AB372072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pot_sequences.
 XX
 SQ Sequence 1480 AA;
 Query Match 72.0%; Score 36; DB 22; Length 1480;
 Best Local Similarity 75.0%; Pred. No. 1.4e+03;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 STQTPPPNG 8
 DB 999 SNQSPPPNG 1006
 RESULT 11
 ID AAU33195 standard; Protein; 2153 AA.
 AAU33195
 XX
 AC AAU33195;
 XX 18-DEC-2001 (first entry)
 DT Novel human secreted protein #3686.
 DE Human; vaccination; gene therapy; nutritional supplement;
 XX stem cell proliferation; haematopoiesis; nerve tissue regeneration;
 KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
 XX Homo sapiens.
 OS WO200179449-A2.
 XX 25-OCT-2001.
 PD 16-APR-2001; 2001WO-US08656.
 PF 18-APR-2000; 2000US-0552929.
 PR 26-JAN-2001; 2001US-0770160.
 XX (HYSE-) HYSEQ INC.
 PA Tang YT, Liu C, Drmanac RT;
 XX WPI; 2001-611725/70.
 DR Nucleic acids encoding a range of human polypeptides, useful in genetic
 PT vaccination, testing and therapy -
 PT
 XX Claim 20; Page 727-728; 765pp; English.
 XX The invention relates to novel human secreted polypeptides. The
 CC polypeptides and antibodies to the polypeptides are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of polypeptide. The polypeptides are also useful for
 CC identifying agents (agonists and antagonists) that bind to them. Cells
 CC expressing the proteins are useful for identifying a therapeutic agent
 CC for use in treatment of a pathology related to aberrant expression or
 CC physiological interactions of the polypeptide. Vectors comprising
 CC the nucleic acids encoding the polypeptides and cells genetically

SQ Sequence 322 AA;
Query Match 72.0%; Score 36; DB 22; Length 322;
Best Local Similarity 85.7%; Pred. No. 3.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 3 QTPPNGR 9
DB 314 QTPPDGR 320
RESULT 7
ID AAY44944 standard; Protein; 656 AA.
XX AC AAY44944;
XX DT 23-MAY-2000 (first entry)
XX DE Wheat sulphate permealase-1.
XX KW Sulphate Permealase; sulphate assimilation protein; wheat; probe;
XX KW mapping; marker; plant breeding; chimeric gene; transgenic plant;
XX KW antibody; screen.
XX OS Triticum aestivum.
XX PN WO200004154-A2.
XX PD 27-JAN-2000.
XX PF 13-JUL-1999; 99WO-US15810.
XX PR 14-JUL-1998; 98US-0092833.
XX PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX PI Allen SM, Falco SC, Thorpe CJ;
XX WPI; 2000-195025/17.
XX DR N-PSDB; AA44944.
XX PT Nucleic acid fragments encoding sulfate assimilation proteins in plants
XX PT and seeds useful as probes for isolating cDNAs and genes encoding
XX PT homologous proteins, in producing transgenic plants -
XX PS Claim 6; Page 56-58; 79pp; English.
XX CC The present amino acid sequence is the wheat sulphate permealase, a
XX CC sulphate assimilation protein. This sequence is obtained from wki1
XX CC library, clone wki.pk0028.e1, derived from wheat seedlings, 1 hour
XX CC after treatment with fungicide. It has 71% sequence identity to Hordeum
XX CC vulgare sulphate permealase (gi 1217967).
XX CC This sequence is used as a probe to isolate other plant sulphate
XX CC assimilation proteins, for genetic and physical mapping of related genes
XX CC and as markers of traits linked to the gene. This is useful for plant
XX CC breeding and to construct chimeric genes, used to create transgenic
XX CC plants with altered levels of sulphate permealase. The sulphate permealase
XX CC peptides are useful for producing antibodies, that are used to screen
XX CC and isolate cDNA clones.
XX SQ Sequence 656 AA;
Query Match 72.0%; Score 36; DB 21; Length 656;
Best Local Similarity 85.7%; Pred. No. 6.4e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 STQTPPN 7
DB 16 TTQTPPN 22
RESULT 8
ID AAG91907 standard; Protein; 718 AA.
XX AC AAG91907;
XX DT 26-SEP-2001 (first entry)
XX DE C glutamicum protein fragment SEQ ID NO: 5661.
XX KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
XX KW organic acid synthesis.
XX OS Corynebacterium glutamicum.
XX PN EP1108790-A2.
XX PD 20-JUN-2001.
XX PF 18-DEC-2000; 2000EP-0127688.
XX PR 16-DEC-1999; 99JP-0377484.
XX PR 07-APR-2000; 2000JP-0159162.
XX PR 03-AUG-2000; 2000JP-0280988.
XX PA (KYOW) KYOWA HAKKO KOGYO KK.
XX PN Nakagawa S, Mizoguchi H, Ando S, Hayaishi M, Ochiai K, Yokoi H;
XX PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX WPI; 2001-376931/40.
XX DR N-PSDB; AA67126.
XX PT Novel polynucleotides derived from Coryneform bacteria, for identifying
XX PT mutation point of a gene, measuring expression of a gene, analysing
XX PT expression profile or pattern of a gene and identifying homologous gene
XX PS Claim 17; SEQ ID NO: 5661; 246pp + Sequence Listing; English.
XX CC The present invention provides a number of nucleotide and protein
XX CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
XX CC are useful for identifying the mutation point of a gene derived from a
XX CC mutant of coryneform bacterium, measuring expression amount and
XX CC analysing the expression profile or expression pattern of a gene derived
XX CC from Coryneform bacterium, and identifying a homologue of a gene derived
XX CC from Coryneform bacterium. Coryneform bacteria are useful for producing
XX CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
XX CC particularly L-lysine. The present sequence is a protein described
XX CC in the exemplification of the invention.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from the
XX CC European Patent Office.
XX SQ Sequence 718 AA;
Query Match 72.0%; Score 36; DB 22; Length 718;
Best Local Similarity 85.7%; Pred. No. 7e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 3 QTPPNGR 9
DB 314 QTPPDGR 320
RESULT 9
ID ABP56416 standard; Protein; 978 AA.
XX AC ABP56416;
XX DT 12-MAR-2003 (first entry)
XX DE Thermus thermophilus DNA repair enzyme TRCF protein SEQ ID NO:8.

```
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149923.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150565.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151309.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158023.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.

PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 84.8%; Score 39; DB 21; Length 163;
Best Local Similarity 88.9%; Pred. No. 33;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 STEGEGSGR 9
DB 121 STYGEGR 129

RESULT 9
AAG53082
ID AAG53082 standard; Protein; 163 AA.
XX AC AAG53082;
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 57548.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; Genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
XX PR 23-MAR-1999; 99US-0125788.
XX PR 25-MAR-1999; 99US-0126264.
XX PR 29-MAR-1999; 99US-0126785.
XX PR 01-APR-1999; 99US-0127462.
XX PR 06-APR-1999; 99US-0128234.
XX PR 08-APR-1999; 99US-0128714.
XX PR 16-APR-1999; 99US-0128845.
XX PR 19-APR-1999; 99US-0130077.
XX PR 21-APR-1999; 99US-0130449.
XX PR 23-APR-1999; 99US-0130510.
XX PR 23-APR-1999; 99US-0130891.
XX PR 30-APR-1999; 99US-0131449.
XX PR 30-APR-1999; 99US-0132048.
XX PR 30-APR-1999; 99US-0132407.
XX PR 04-MAY-1999; 99US-0132484.
XX PR 05-MAY-1999; 99US-0132485.
XX PR 06-MAY-1999; 99US-0132486.
XX PR 06-MAY-1999; 99US-0132487.
XX PR 07-MAY-1999; 99US-0132863.
XX PR 11-MAY-1999; 99US-0134256.
```

PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161932.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 84.8%; Score 39; DB 21; Length 163;

Best Local Similarity 88.9%; Pred. No. 33;

Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 STEGEGSGR 9

DB 121 STYEGSGR 129

RESULT 10

AAG11546
ID AAG11546 standard; Protein; 174 AA.

AC AAG11546;

XX 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 10303.

XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; Genetic mapping; gene expression control; promoter;
KW termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132867.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 28-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.

PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139753.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139859.
 PR 23-JUN-1999; 99US-0140352.
 PR 23-JUN-1999; 99US-0140353.
 PR 24-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 28-JUL-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.

PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149358.
 PR 17-AUG-1999; 99US-0149317.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157855.
 PR 07-OCT-1999; 99US-0158023.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159283.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161922.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 84.8%; Score 39; DB 21; Length 174;
 Best Local Similarity 88.9%; Pred. No. 35;

XX Sequence 174 AA;
SQ Query Match 84.8%; Score 39; DB 24; Length 174;
Best Local Similarity 88.9%; Pred. No. 35;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 STEGEGSGR 9
DB 132 STYEGSGR 140

RESULT 14
ID AAR60796 standard; Protein; 1071 AA.
AC AAR60796;
XX 13-JUL-1995 (first entry)
DT 13-JUL-1995 (first entry)
DE Rice sucrose phosphate synthetic enzyme.
XX Sucrose phosphate; synthetic enzyme;
KW Increased sugar and starch yields.
XX Oryza sativa.
OS
XX JP06277068-A.
XX 04-OCT-1994.
XX 27-MAR-1993; 93JP-0092520.
XX 27-MAR-1993; 93JP-0092520.
XX (MITK) MITSUI TOATSU CHEM INC.
XX WPI; 1994-353749/44.
DR N-PSDB; AAQ73703.
XX Gene for sucrose phosphate synthetic enzyme of rice - useful for
PT increasing sugar and starch yield of plants
XX Claim 1; Page 5; 18pp; Japanese.
XX AAQ73703 encodes AAR60796 rice sucrose phosphate synthetic enzyme,
CC this gene may be cloned into other edible plants to improve
CC their synthesis of sucrose and starch.
XX Sequence 1071 AA;
SQ Query Match 82.6%; Score 38; DB 15; Length 1071;
Best Local Similarity 87.5%; Pred. No. 3.3e+02;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 STEGEGSG 8
DB 254 STDGEGSG 261

RESULT 15
ID AAU01686 standard; Protein; 25 AA.
XX AAU01686;
XX 18-JUL-2001 (first entry)
DT Human secreted protein encoded by gene #28.
DE Human secreted protein; diagnosis; autoimmune disease;
XX rheumatoid arthritis; hyperproliferative disorder; neoplasm; sunburn;
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;

KW cerebral ischaemia; angiogenesis; nervous system disorder; skin aging;
KW Alzheimer's disease; infection; ocular disorder; corneal infection;
KW wound healing; epithelial cell proliferation; chemotaxis; preservative;
XX organ transplantation; tissue regeneration; food additive.
OS Homo sapiens.
XX WC200123409-A2.
XX 05-APR-2001.
XX 26-SEP-2000; 2000WO-US26371.
XX 27-SEP-1999; 99US-0155804.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Ruben SM, Komatsoulis GA;
XX WPI; 2001-266139/27.
DR N-PSDB; AAS02547.
XX Nucleic acids encoding 38 human secreted polypeptides, useful for
PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's
XX disease and diabetic retinopathy -
XX Disclosure; Page 47; 48pp; English.
XX AAU01641-AAU01698 represent human secreted protein amino acid, and
CC related amino acid sequences of the invention. The human secreted protein
CC sequences are used to prevent, treat or ameliorate a medical condition in
CC e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep.
CC They are also used in diagnosing a pathological condition or
CC susceptibility to a pathological condition. The antibodies to human
CC secreted proteins can also be used in alleviating symptoms associated
CC with the disorders and in diagnostic immunoassays e.g. radioimmunoassays
CC or enzyme linked immunosorbent assays (ELISA). Disorders which are
CC diagnosed or treated include autoimmune diseases e.g. rheumatoid
CC arthritis, hyperproliferative disorders e.g. neoplasms of the breast or
CC liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular
CC disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders
CC e.g. Alzheimer's disease, infections caused by bacteria, viruses and
CC fungi and ocular disorders e.g. corneal infection. The polypeptides can
CC also be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities.
XX SQ Sequence 25 AA;
Query Match 80.4%; Score 37; DB 22; Length 25;
Best Local Similarity 87.5%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TEGEGSGR 9
DB 13 TEGEGSGR 20

Search completed: December 11, 2003, 18:24:55
Job time : 35 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 18:21:01 ; Search time 8.22222 Seconds
(without alignments)
46.313 Million cell updates/sec

Title: US-10-014-658-6

Perfect score: 46

Sequence: 1 STEGEGSGR 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgm2_6/ptodata/2/iaa/5A_COMB.pdp.*
- 2: /cgm2_6/ptodata/2/iaa/5B_COMB.pdp.*
- 3: /cgm2_6/ptodata/2/iaa/6A_COMB.pdp.*
- 4: /cgm2_6/ptodata/2/iaa/6B_COMB.pdp.*
- 5: /cgm2_6/ptodata/2/iaa/PCTUS_COMB.pdp.*
- 6: /cgm2_6/ptodata/2/iaa/backfiles1.pdp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	82.6	1084	US-09-394-272-9	Sequence 9, Appli
2	36	78.3	71	US-08-280-443-18	Sequence 18, Appl
3	36	78.3	71	US-08-457-459-18	Sequence 18, Appl
4	36	78.3	71	US-08-555-678-18	Sequence 18, Appl
5	36	78.3	71	PCT-US95-02275-18	Sequence 18, Appl
6	36	78.3	72	US-09-286-959B-18	Sequence 18, Appl
7	35	76.1	38	US-08-857-046A-21	Sequence 21, Appl
8	34	73.9	166	US-09-252-991A-20150	Sequence 20150, A
9	34	73.9	201	US-08-220-828-116	Sequence 116, App
10	34	73.9	246	US-08-252-991A-30011	Sequence 30011, A
11	34	73.9	394	US-08-673-814-6	Sequence 6, Appli
12	34	73.9	394	US-09-115-824-6	Sequence 6, Appli
13	34	73.9	473	US-08-103-739B-2	Sequence 2, Appli
14	34	73.9	473	US-08-474-404-2	Sequence 2, Appli
15	34	73.9	473	US-08-485-845-2	Sequence 2, Appli
16	34	73.9	473	US-08-482-714-2	Sequence 2, Appli
17	34	73.9	473	US-08-211-416-2	Sequence 2, Appli
18	34	73.9	473	US-09-059-958-2	Sequence 2, Appli
19	34	73.9	846	US-08-252-991A-23780	Sequence 23780, A
20	33	71.7	93	US-08-902-623-43	Sequence 43, Appl
21	33	71.7	227	US-08-272-882D-2	Sequence 2, Appli
22	33	71.7	227	PCT-US92-05532-2	Sequence 2, Appli
23	33	71.7	310	US-08-489-537A-74	Sequence 74, Appl
24	33	71.7	517	US-08-252-991A-31115	Sequence 31115, A
25	33	71.7	595	US-08-252-991A-25815	Sequence 25815, A
26	33	71.7	641	US-08-167-206-4	Sequence 4, Appli
27	33	71.7	790	US-08-359-705B-9	Sequence 9, Appli

28 33 71.7 790 2 US-08-286-846A-9 Sequence 9, Appli
29 33 71.7 790 2 US-08-457-880A-9 Sequence 9, Appli
30 33 71.7 790 3 US-08-444-622A-9 Sequence 9, Appli
31 33 71.7 790 3 US-08-942-562-9 Sequence 9, Appli
32 33 71.7 790 3 US-09-156-923-9 Sequence 9, Appli
33 33 71.7 814 1 US-08-286-305A-3 Sequence 3, Appli
34 33 71.7 814 2 US-08-441-104A-3 Sequence 3, Appli
35 33 71.7 814 2 US-08-440-816A-3 Sequence 3, Appli
36 33 71.7 814 3 US-09-417-381A-3 Sequence 3, Appli
37 33 71.7 882 3 US-09-413-814-78 Sequence 78, Appli
38 32 69.6 91 4 US-09-732-210-1276 Sequence 1276, Ap
39 32 69.6 143 2 US-08-942-819-14 Sequence 14, Appl
40 32 69.6 143 4 US-09-522-955A-22 Sequence 22, Appl
41 32 69.6 174 3 US-09-097-889-24 Sequence 24, Appl
42 32 69.6 174 4 US-09-098-079-24 Sequence 24, Appl
43 32 69.6 228 4 US-09-252-991A-26379 Sequence 26379, A
44 32 69.6 384 4 US-09-252-991A-31756 Sequence 31756, A
45 32 69.6 403 2 US-08-474-379C-14 Sequence 14, Appl

ALIGNMENTS

RESULT 1

US-09-394-272-9
; Sequence 9, Application US/09394272
; Patent No. 6472588
; GENERAL INFORMATION:
; APPLICANT: Haigler, Candace H.
; APPLICANT: Holaday, A. Scott
; TITLE OF INVENTION: TRANSGENIC FIBER PRODUCING PLANTS WITH INCREASED
; TITLE OF INVENTION: EXPRESSION OF SUCROSE PHOSPHATE SYNTHASE
; FILE REFERENCE: 201304/1000
; CURRENT APPLICATION NUMBER: US/09/394,272
; CURRENT FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1084
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-394-272-9

Query Match 82.6%; Score 38; DB 4; Length 1084;
Best Local Similarity 87.5%; Pred. No. 1.4e+02;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 STEGEGSG 8
DB 268 STDGEGSG 275

RESULT 2

US-08-280-443-18
; Sequence 18, Application US/08280443
; Patent No. 5643778
; GENERAL INFORMATION:
; APPLICANT: Nishikura, Kazuko
; TITLE OF INVENTION: RNA Editing Enzyme and Methods of Use
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, P.O. Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

PCT-US95-02275-18
; Sequence 18, Application PC/TUS9502275
; GENERAL INFORMATION:
; APPLICANT: Wistar Institute of Anatomy & Biology
; TITLE OF INVENTION: RNA Editing Enzyme and Methods of Use
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, P.O. Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/02275
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/280,443
; FILING DATE: 25-JUL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/197,794
; FILING DATE: 17-FEB-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: WST49BPCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9206
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 71 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
PCT-US95-02275-18

Query Match 78.3%; Score 36; DB 5; Length 71;
Best Local Similarity 75.0%; Pred. No. 21;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 TEGEGSGR 9
Db 45 TEGEGNGK 52

RESULT 6
US-09-286-959B-18
; Sequence 18, Application US/09286959B
; Patent No. 6300131
; GENERAL INFORMATION:
; APPLICANT: Johns Hopkins University
; APPLICANT: Greider, Carol W.
; APPLICANT: Le, Siyuan
; TITLE OF INVENTION: TELOMERASE-ASSOCIATED PROTEINS
; FILE REFERENCE: 07265/157001
; CURRENT APPLICATION NUMBER: US/09/286,959B
; CURRENT FILING DATE: 1999-04-06
; PRIOR APPLICATION NUMBER: 60/080,783
; PRIOR FILING DATE: 1998-04-06
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 72
; TYPE: PRT

; ORGANISM: Drosophila melanogaster
; FEATURE:
; NAME/KEY: BINDING
; LOCATION: (0)...(0)
US-09-286-959B-18

Query Match 78.3%; Score 36; DB 4; Length 72;
Best Local Similarity 75.0%; Pred. No. 21;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 TEGEGSGR 9
Db 46 TEGEGNGK 53

RESULT 7
US-08-857-046A-21
; Sequence 21, Application US/08857046A
; Patent No. 6361938
; GENERAL INFORMATION:
; APPLICANT: O'Mahony, Daniel J
; APPLICANT: Alveez, Vernon L
; APPLICANT: Seveso, Michela
; TITLE OF INVENTION: Peptides Which Enhance Transport Across
; Tissues and Methods of Identifying and Using the Same
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mary L. Severson, Ph.D., Esq.
; STREET: 1300 Gould Drive
; CITY: Gainesville
; STATE: GA
; COUNTRY: USA
; ZIP: 30504
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/857,046A
; FILING DATE: 15-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/006461
; FILING DATE: 10-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IE 950864
; FILING DATE: 10-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/746,411
; FILING DATE: 08-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/IE96/00073
; FILING DATE: 11-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/IE96/00072
; FILING DATE: 11-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Severson, Mary L
; REGISTRATION NUMBER: 34,927
; REFERENCE/DOCKET NUMBER: 97,1061.US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 770 534-8239
; TELEFAX: 770 534-8247
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 38 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-857-046A-21

Query Match 76.1%; Score 35; DB 4; Length 38;

Best Local Similarity 75.0%; Pred. No. 2.4e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 STEGEGSG 8

Db 111 SSEGQSG 118

RESULT 12

US-09-115-824-6
; Sequence 6, Application US/09115824
; Patent No. 6326001

GENERAL INFORMATION:

APPLICANT: Inzana, Thomas J.

ADDRESS: Ward, Christine

TITLE OF INVENTION: RECOMBINANT VACCINE FOR DISEASES CAUSED
BY ENCAPSULATED ORGANISMS

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: Whitham, Curtis, Whitham & McGinn

STREET: 11800 Sunrise Valley Dr., Suite 900

CITY: Reston

STATE: VA

COUNTRY: USA

ZIP: 20191

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/115,824

FILING DATE: 15-Jul-1998

CLASSIFICATION: unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/673,814

FILING DATE: 27-JUN-1996

ATTORNEY/AGENT INFORMATION:

NAME: Whitham, Michael E.

REFERENCE/DOCKET NUMBER: VTIP 95-067

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-391-2510

TELEFAX: 703-391-9035

TELEX: 283072

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 394 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 6:

US-09-115-824-6

Query Match 73.9%; Score 34; DB 4; Length 394;

Best Local Similarity 75.0%; Pred. No. 2.4e+02;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 STEGEGSG 8

Db 111 SSEGQSG 118

RESULT 13

US-08-103-739B-2

; Sequence 2, Application US/08103739B

; Patent No. 5478369

GENERAL INFORMATION:

APPLICANT: ALBERTSEN, Marc C.

APPLICANT: BEACH, Larry R.

APPLICANT: HOWARD, John A.

APPLICANT: HUFFMAN, Gary A.

TITLE OF INVENTION: DNA Sequences Mediating Male Fertility

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pioneer Hi-Bred International, Inc.

STREET: 700 Capital Square, 400 Locust Street

CITY: Des Moines

STATE: Iowa

COUNTRY: U.S.

ZIP: 50309

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/103,739B

FILING DATE: 02-AUG-1993

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/537,183

FILING DATE: 12-JUN-1990

ATTORNEY/AGENT INFORMATION:

NAME: Sweeney, Patricia A.

REGISTRATION NUMBER: 32,733

REFERENCE/DOCKET NUMBER: 0125R2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (515) 248 4897

TELEFAX: (515) 248-4844

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 473 amino acids

TYPE: amino acid

TOPOLOGY: linear

US-08-103-739B-2

Query Match 73.9%; Score 34; DB 1; Length 473;

Best Local Similarity 85.7%; Pred. No. 2.8e+02;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 EGEGR 9

Db 238 EGEGR 244

RESULT 14

US-08-474-404-2

; Sequence 2, Application US/08474404

; Patent No. 5824524

GENERAL INFORMATION:

APPLICANT: ALBERTSEN, Marc C.

APPLICANT: BEACH, Larry R.

APPLICANT: HOWARD, John A.

APPLICANT: HUFFMAN, Gary A.

TITLE OF INVENTION: Nucleotide Sequences Mediating Fertility and Method of Using

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pioneer Hi-Bred International, Inc.

STREET: 700 Capital Square, 400 Locust Street

CITY: Des Moines

STATE: Iowa

COUNTRY: U.S.

ZIP: 50309

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/474,404

FILING DATE: 07-JUN-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/103,739

FILING DATE: 02-AUG-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/537,183

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 18:21:27 ; Search time 49.3333 Seconds
(without alignments)
33.929 Million cell updates/sec

Title: US-10-014-658-6

Perfect score: 46

Sequence: 1 STEGEGSGR 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 684280 seqs, 185983659 residues

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*

- 1: /cgn2_6/ptodata/1/pubaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubaa/FCR_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubaa/PTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39	84.8	174	11	US-09-819-142-8
2	39	84.8	174	11	US-09-819-142-8
3	38	82.6	1084	15	US-09-934-455-38
4	35	76.1	38	12	US-10-217-700-9
5	35	76.1	609	9	US-10-116-275-5
6	34	73.9	127	9	US-09-815-242-10749
7	34	73.9	201	9	US-09-815-242-10432
8	34	73.9	278	15	US-09-220-920-116
9	34	73.9	340	15	US-10-156-761-10476
10	34	73.9	411	10	US-10-083-357-1337
11	34	73.9	839	12	US-09-738-626-6175
12	34	73.9	839	12	US-10-237-496-54
13	34	73.9	839	12	US-10-242-074-54
14	34	73.9	839	12	US-10-242-505-54
15	34	73.9	839	12	US-10-242-574-54
16	34	73.9	839	12	US-10-243-261-54

15	34	73.9	839	12	US-10-243-282-54	Sequence 54, Appl
17	34	73.9	839	12	US-10-243-402-54	Sequence 54, Appl
18	34	73.9	839	12	US-10-243-431-54	Sequence 54, Appl
19	34	73.9	839	12	US-10-243-464-54	Sequence 36, Appl
20	34	73.9	839	12	US-10-210-951-36	Sequence 54, Appl
21	34	73.9	839	12	US-10-244-972-54	Sequence 54, Appl
22	34	73.9	839	12	US-10-197-942-54	Sequence 36, Appl
23	34	73.9	839	12	US-10-211-884-36	Sequence 54, Appl
24	34	73.9	839	12	US-10-238-196-54	Sequence 54, Appl
25	34	73.9	839	12	US-10-245-103-54	Sequence 54, Appl
26	34	73.9	839	15	US-10-245-107-54	Sequence 54, Appl
27	34	73.9	839	15	US-10-245-143-54	Sequence 54, Appl
28	34	73.9	839	15	US-10-245-171-54	Sequence 54, Appl
29	34	73.9	839	15	US-10-245-851-54	Sequence 54, Appl
30	34	73.9	839	15	US-10-245-883-54	Sequence 54, Appl
31	34	73.9	839	15	US-10-237-535-54	Sequence 54, Appl
32	34	73.9	839	15	US-10-238-183-54	Sequence 54, Appl
33	34	73.9	839	15	US-10-238-283-54	Sequence 54, Appl
34	34	73.9	839	15	US-10-238-370-54	Sequence 54, Appl
35	34	73.9	839	15	US-10-245-055-54	Sequence 54, Appl
36	34	73.9	839	15	US-10-245-147-54	Sequence 54, Appl
37	34	73.9	839	15	US-10-245-730-54	Sequence 54, Appl
38	34	73.9	839	15	US-10-245-739-54	Sequence 54, Appl
39	34	73.9	839	15	US-10-246-210-54	Sequence 54, Appl
40	34	73.9	839	15	US-10-239-136-54	Sequence 54, Appl
41	34	73.9	839	15	US-10-243-024-54	Sequence 54, Appl
42	34	73.9	839	15	US-10-243-409-54	Sequence 54, Appl
43	34	73.9	839	15	US-10-245-621-54	Sequence 54, Appl
44	34	73.9	839	15	US-10-245-880-54	Sequence 54, Appl
45	34	73.9	839	15	US-10-245-880-54	Sequence 54, Appl

ALIGNMENTS

RESULT 1

US-09-819-142-8
; Sequence 8, Application US/09819142
; Publication No. US20030041356A1
; GENERAL INFORMATION:
; APPLICANT: Reuber, Lynne
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Heard, Jacqueline
; APPLICANT: Adam, Luc
; TITLE OF INVENTION: Methods for Modifying Flowering Phenotypes
; FILE REFERENCE: MEI-0033
; CURRENT APPLICATION NUMBER: US/09/819,142
; CURRENT FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent version 3.0
; SEQ ID NO 8
; LENGTH: 174
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-819-142-8

Query Match 84.8%; Score 39; DB 11; Length 174;
Best Local Similarity 88.9%; Pred. No. 47;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 STEGEGSGR 9
|||
Db 132 STYEGSGR 140

RESULT 2

US-09-934-455-38
; Sequence 38, Application US/09934455
; Publication No. US20030121070A1
; GENERAL INFORMATION:
; APPLICANT: Adam, Luc

Best Local Similarity 100.0%; Pred. No. 6.8e+02; Indels 0; Gaps 0;
Matches 7; Conservative 0; Mismatches 0;

Qy 1 STEGEGS 7
Db 586 STEGEGS 592

RESULT 6

US-09-815-242-10432
; Sequence 10432, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10432
; LENGTH: 127
; TYPE: PRT
; ORGANISM: Escherichia coli

US-09-815-242-10432

Query Match 73.9%; Score 34; DB 9; Length 127;
Best Local Similarity 66.7%; Pred. No. 2.2e+02; Indels 1; Gaps 0;
Matches 6; Conservative 2; Mismatches 1;

Qy 1 STEGEGS 9
Db 4 NTEAQSGR 12

RESULT 7

US-09-220-920-116
; Sequence 116, Application US/09220920
; Patent No. US2002002269A1
; GENERAL INFORMATION:
; APPLICANT: Milbrandt, Jeffrey D.
; APPLICANT: Baloh, Robert H.
; TITLE OF INVENTION: Artemin, A No. US2002002269A1e1 Neurotrophic Factor
; FILE REFERENCE: 6029-7996
; CURRENT APPLICATION NUMBER: US/09/220,920
; CURRENT FILING DATE: 1998-12-24
; EARLIER APPLICATION NUMBER: 09/163,283
; EARLIER FILING DATE: 1998-09-29
; EARLIER APPLICATION NUMBER: 60/108,148
; EARLIER FILING DATE: 1998-11-12
; EARLIER APPLICATION NUMBER: 09/218,698

; EARLIER FILING DATE: 1998-12-22
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 116
; LENGTH: 201
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-220-920-116

Query Match 73.9%; Score 34; DB 9; Length 201;
Best Local Similarity 85.7%; Pred. No. 3.4e+02; Indels 0; Gaps 0;
Matches 6; Conservative 1; Mismatches 0;

Qy 3 EGECSGR 9
Db 1 EGECSGR 7

RESULT 8

US-10-156-761-10476
; Sequence 10476, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 10476
; LENGTH: 278
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-10476

Query Match 73.9%; Score 34; DB 15; Length 278;
Best Local Similarity 66.7%; Pred. No. 4.7e+02; Indels 1; Gaps 0;
Matches 6; Conservative 2; Mismatches 1;

Qy 1 STEGEGS 9
Db 138 STDGSGSQ 146

RESULT 9

US-10-083-357-1337
; Sequence 1337, Application US/10083357
; Publication No. US20030054370A1
; GENERAL INFORMATION:
; APPLICANT: Qianlong Zeng et al.
; TITLE OF INVENTION: Systemic Discovery of New Genes
; FILE REFERENCE: 032796-090
; CURRENT APPLICATION NUMBER: US/10/083,357
; CURRENT FILING DATE: 2002-02-27
; NUMBER OF SEQ ID NOS: 1346
; SEQ ID NO 1337
; LENGTH: 340
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-083-357-1337

Query Match 73.9%; Score 34; DB 15; Length 340;
Best Local Similarity 85.7%; Pred. No. 5.7e+02;

```
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
; PRIOR FILING DATE: 1998-06-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 54
; LENGTH: 839
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-242-074-54

Query Match      73.9%; Score 34; DB 12; Length 839;
Best Local Similarity 85.7%; Pred. NO. 1.3e+03;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      3 EGEGR 9
      |||:|
Db      818 EGEGR 824

RESULT 14
US-10-242-574-54
; Sequence 54, Application US/10242574
; Publication No. US20030138899A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Eaton, Dan
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Goddard, Audrey
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Phillippe
; APPLICANT: Watanabe, Colin
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; APPLICANT: Fong, Sherman
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3630R1C20
; CURRENT APPLICATION NUMBER: US/10/242,574
; CURRENT FILING DATE: 2002-09-11
; PRIOR APPLICATION NUMBER: 10/197942
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/059114
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/065027
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/086478
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 54
; LENGTH: 839
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-242-574-54

Query Match      73.9%; Score 34; DB 12; Length 839;
Best Local Similarity 85.7%; Pred. NO. 1.3e+03;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      3 EGEGR 9
      |||:|
Db      818 EGEGR 824

RESULT 15
US-10-243-261-54
; Sequence 54, Application US/10243261
; Publication No. US20030138900A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Eaton, Dan
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 18:19:54 ; Search time 8.33333 Seconds
(without alignments)
103.862 Million cell updates/sec

Title: US-10-014-658-6

Perfect score: 46

Sequence: 1 STEGEGSGR 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_76:*

1: Pirl1:*

2: Pirl2:*

3: Pirl3:*

4: Pirl4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39	84.8	167	2	T52599
2	39	84.8	174	2	T52600
3	38	82.6	1084	2	T04103
4	36	78.3	212	2	F82581
5	36	78.3	697	2	T34704
6	36	78.3	1026	1	A40315
7	35	76.1	1527	2	JE0336
8	34	73.9	127	1	VXECSE
9	34	73.9	127	2	H91241
10	34	73.9	127	2	E8089
11	34	73.9	127	2	AC0934
12	34	73.9	203	2	AC1393
13	34	73.9	246	2	B83418
14	34	73.9	259	1	YXBYT
15	34	73.9	305	2	AI2079
16	34	73.9	394	1	BWHIXD
17	34	73.9	402	2	T14438
18	34	73.9	502	2	F31751
19	34	73.9	795	2	T36588
20	34	73.9	2248	1	D42088
21	34	73.9	4957	2	T03454
22	34	73.9	5262	2	T03454
23	33	71.7	118	2	I57506
24	33	71.7	118	2	I77386
25	33	71.7	142	2	T25224
26	33	71.7	196	2	S08677
27	33	71.7	227	2	JH0571
28	33	71.7	295	2	JU0043
29	33	71.7	306	2	AG3542

30	33	71.7	349	2	T23931
31	33	71.7	422	1	BMRT2Y
32	33	71.7	422	2	A55417
33	33	71.7	480	2	D86674
34	33	71.7	484	2	F83330
35	33	71.7	509	2	S18872
36	33	71.7	533	1	O4RTV3
37	33	71.7	535	2	A54155
38	33	71.7	536	2	S71332
39	33	71.7	540	1	OYHUCR
40	33	71.7	710	2	E83360
41	33	71.7	790	1	TVHUTT
42	33	71.7	791	2	F81056
43	33	71.7	799	1	TVRTTB
44	33	71.7	862	2	S30911
45	33	71.7	937	2	S58135

ALIGNMENTS

RESULT 1

T52599
squamosa promoter binding protein 4 [imported] - Arabidopsis thaliana (fragment)
C/Species: Arabidopsis thaliana (mouse-ear cross)
C/Date: 24-Oct-2000 #sequence_revision 24-Oct-2000 #text_change 24-Oct-2000
C/Accession: T52599
R/Cardon, G.H.; Hoehmann, S.; Klein, J.; Nettesheim, K.; Saedler, H.; Huijser, P.
Gene 237, 91-104, 1999
A/Title: Molecular Characterisation of the Arabidopsis SBP-box genes.
A/Reference number: 225236; MUID:99453765; PMID:10524240
A/Accession: T52599
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-167 <CAR>
A/Cross-references: EMBL:AJ011632; PIDN:CAB56584.1
A/Experimental source: cultivar Landsberg erecta; inflorescence
C/Genetics:
A/Gene: spl4
C/Keywords: DNA binding; transcription factor

Query Match 84.8%; Score 39; DB 2; Length 167;
Best Local Similarity 88.9%; Pred. No. 5.8;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	STEGEGSGR	9
DB	125	STYEGSGR	133

RESULT 2

T52600
squamosa promoter binding protein 4 [imported] - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cross)
C/Date: 24-Oct-2000 #sequence_revision 24-Oct-2000 #text_change 24-Oct-2000
C/Accession: T52600
R/Cardon, G.H.; Hoehmann, S.; Klein, J.; Nettesheim, K.; Saedler, H.; Huijser, P.
Gene 237, 91-104, 1999
A/Title: Molecular Characterisation of the Arabidopsis SBP-box genes.
A/Reference number: 225236; MUID:99453765; PMID:10524240
A/Accession: T52600
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-174 <CAR>
A/Cross-references: EMBL:AJ011631; PIDN:CAB56583.1
A/Experimental source: cultivar Landsberg erecta; inflorescence
C/Genetics:
A/Gene: spl4
C/Keywords: DNA binding; transcription factor

Query Match 84.8%; Score 39; DB 2; Length 174;
Best Local Similarity 88.9%; Pred. No. 6.1;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

hypothetical prote
synaptotagmin II -
synaptotagmin II -
sensor protein kin
RND divalent metal
legumin-like stora
vitamin D3 25-mono
natriuretic peptid
natriuretic peptid
conserved hypothet
nerve growth facto
hemoglobin recepto
nerve growth facto
SMP2 protein - yea
hyphally regulated

C;Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 04-Jun-1999
C;Accession: JE0336
R;Uchiyama, T.; Hinochita, E.; Haga, S.; Nakamura, T.; Tanaka, T.; Toh, S.; Furukawa, M.;
Biochem. Biophys. Res. Commun. 252, 103-110, 1998
A;Title: Isolation of a novel human canalicular multispecific organic anion transporter,
t.
C;Reference number: JE0336; MUID:99032812; PMID:9813153
A;Accession: JE0336
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1527 <UCH>
A;Cross-references: GB:AF083552
C;Superfamily: human multidrug resistance protein cMOAT2; ATP-binding cassette homology
C;Keywords: ATP
F;1306-1499/Domain: ATP-binding cassette homology <ABC2>

Query Match 76.1%; Score 35; DB 2; Length 1527;
Best Local Similarity 66.7%; Pred. No. 2.9e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 STEGEGSGR 9
:|||||
Db 911 SDDGEGQR 919
:|||||

RESULT 8
VXECSE
preprotein translocase secE chain - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 01-Mar-2002
C;Accession: A35139; A32873; H5204
R;Downing, W.L.; Sullivan, S.L.; Gottesman, M.E.; Dennis, P.P.
J. Bacteriol. 172, 1621-1627, 1990
A;Title: Sequence and transcriptional pattern of the essential Escherichia coli secE-nu
A;Reference number: A35139; MUID:90170882; PMID:2137819
A;Accession: A35139
A;Molecule type: DNA
A;Residues: 1-127 <DOW>
A;Cross-references: GB:M30610; NID:g147798; PIDN:AA24621.1; PID:g147800
R;Schatz, P.J.; Riggs, P.D.; Jacq, A.; Fath, M.J.; Beckwith, J.
Genes Dev. 3, 1035-1044, 1989
A;Title: The secE gene encodes an integral membrane protein required for protein export
A;Reference number: A32873; MUID:89378734; PMID:2673920
A;Accession: A32873
A;Molecule type: DNA
A;Residues: 1-127 <SCH>
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: H5204
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-127 <BLAT>
A;Cross-references: GB:AE000472; GB:U00096; NID:g2367333; PIDN:AAC76955.1; PID:g1790413;
A;Experimental source: strain K-12, substrain MG1655
C;Comment: This integral inner membrane protein is an essential component of the protein
C;Genetics:
A;Gene: secE
A;Map position: 90 min
C;Superfamily: protein-export protein secE
C;Keywords: inner membrane; protein export

Query Match 73.9%; Score 34; DB 1; Length 127;
Best Local Similarity 66.7%; Pred. No. 36;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 STEGEGSGR 9
:|||||
Db 4 NTEAQSGR 12
:|||||

RESULT 9
H91241
preprotein translocase [imported] - Escherichia coli (strain O157:H7, substrain RIMD 05
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C;Accession: H91241
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: H91241
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-127 <HAY>
A;Cross-references: GB:BA000007; PIDN:BA38327.1; PID:g13364380; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: SEC84904
C;Superfamily: protein-export protein secE

Query Match 73.9%; Score 34; DB 2; Length 127;
Best Local Similarity 66.7%; Pred. No. 36;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 STEGEGSGR 9
:|||||
Db 4 NTEAQSGR 12
:|||||

RESULT 10
E86089
preprotein translocase [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: E86089
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhe
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dialanta, E.; Potamousis, K.; Apodaca
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: E86089
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-127 <STO>
A;Cross-references: GB:AE005174; NID:g12518903; PIDN:AAG59177.1; GSPDB:GN00145; UWGP:Z5
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: secE
C;Superfamily: protein-export protein secE

Query Match 73.9%; Score 34; DB 2; Length 127;
Best Local Similarity 66.7%; Pred. No. 36;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 STEGEGSGR 9
:|||||
Db 4 NTEAQSGR 12
:|||||

RESULT 11
AC0934
preprotein translocase SecE chain [imported] - Salmonella enterica subsp. enterica sero
C;Species: Salmonella enterica subsp. enterica serovar typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C;Accession: AC0934
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher
th, T.; Comerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica sero

Query Match 73.9% Score 34; DB 2; Length 305;
 Best Local Similarity 66.7% Pred. No. 87;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 STEGEGSGR 9
 Db 41 TTEGEGSSK 49

Search completed: December 11, 2003, 18:29:35
 Job time : 12.3333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 16:54:27 ; Search time 4.6667 Seconds
(without alignments)
90.694 Million cell updates/sec

Title: US-10-014-658-6

Perfect score: 46

Sequence: 1 STEGEGSGR 9

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36	78.3	777	1 NTRY AZOBR	P45675 azospirillum
2	36	78.3	1026	1 STRAU DROME	P25159 drosophila
3	35	76.1	1527	1 MRP3 HUMAN	O15438 homo sapien
4	34	73.9	127	1 SECE SCOLI	P16920 escherichia
5	34	73.9	127	1 SECE SALT	Q919K1 salmonella
6	34	73.9	304	1 TYSY YEAST	P06785 saccharomyc
7	34	73.9	394	1 BEXD HAEIN	P22236 haemophilus
8	34	73.9	402	1 CPT2 BRAOL	P52178 brassica ol
9	34	73.9	502	1 KDC2 DROME	P16912 drosophila
10	34	73.9	2248	1 CYAL DROME	P32870 drosophila
11	33	71.7	227	1 SNP2 STRO	P43162 streptomyce
12	33	71.7	295	1 VAL1 CSMV	P18921 chloris str
13	33	71.7	306	1 FMT BRUME	Q8YDB3 bruceella me
14	33	71.7	422	1 SYT2 MOUSE	P46097 mus musculus
15	33	71.7	422	1 SYT2 RAT	P29101 rattus norv
16	33	71.7	533	1 CP27 RAT	P17178 r cytochrom
17	33	71.7	535	1 ANPC RAT	P41740 rattus norv
18	33	71.7	536	1 ANPC MOUSE	P70180 mus musculus
19	33	71.7	541	1 ANPC HUMAN	P17342 homo sapien
20	33	71.7	796	1 TRKA HUMAN	P04629 homo sapien
21	33	71.7	799	1 TRKA RAT	P32573 rattus norv
22	33	71.7	862	1 SMP2 YEAST	P32567 saccharomyc
23	33	71.7	937	1 HYRI CANAL	P46591 candida alb
24	32	69.6	91	1 RS10 METVA	P14039 methanococc
25	32	69.6	174	1 NU6M GORGO	Q34573 gorilla gor
26	32	69.6	174	1 NU6M HUMAN	P03923 homo sapien
27	32	69.6	174	1 NU6M HYLLA	Q95710 hylobates l
28	32	69.6	174	1 NU6M PAPHA	Q922X5 papio hamad
29	32	69.6	174	1 NU6M PONPA	P92700 pongo pygma
30	32	69.6	193	1 ISP2 VITS1	Q9XSD0 vitreoscill
31	32	69.6	280	1 TONE NEIME	P57004 neisseria m
32	32	69.6	283	1 TONE NEIGO	O06432 neisseria g
33	32	69.6	293	1 YP75 MYCTU	Q50646 mycobacteri

ALIGNMENTS

RESULT 1

ID	NTRY	AZOBR	STANDARD;	PRT;	777 AA.
AC	P45675;				
DT	01-NOV-1995	(Rel. 32, Created)			
DT	28-FEB-2003	(Rel. 41, Last sequence update)			
DT	28-FEB-2003	(Rel. 41, Last annotation update)			
DE	Nitrogen regulation protein ntry homolog (EC 2.7.3.-).				
OS	Azospirillum brasilense.				
OC	Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;				
OC	Rhodospirillaceae; Azospirillum.				
OX	NCBI_TaxID=192;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=Sp7 / ATCC 29145;				
RA	Ishida M.L.;				
RL	Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.				
RN	[2]				
RP	SEQUENCE OF 1-300 FROM N.A.				
RC	STRAIN=Sp7 / ATCC 29145;				
RX	MEDLINE:96002420; PubMed:7553451;				
RA	Machado E.B., Yates M.G., Funayama S., Rigo L.U., Steffens M.B.R.,				
RA	Souza E.M., Pedrosa F.O.;				
RT	"The ntrBC genes of Azospirillum brasilense are part of a nifH3-like-				
RT	ntrB-ntrC operon and are negatively regulated."				
RL	Can. J. Microbiol. 41:674-684(1995).				
CC	-1- FUNCTION: MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM NTRY/NTRX				
CC	INVOLVED IN NITROGEN LEVEL CONTROL. PROBABLY ACTIVATES NTRX BY				
CC	PHOSPHORYLATION (BY SIMILARITY).				
CC	-1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).				
CC	-1- SIMILARITY: Contains 1 HAMP domain.				
CC	-1- SIMILARITY: Contains 1 histidine kinase domain.				
CC	-1- SIMILARITY: Contains 1 PAS (PER-ARNT-SIM) dimerization domain.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@sib-sib.ch).				
CC	-----				
DR	EMBL;	Z37984;	CAA86066.2;	--	
DR	PIR;	I39495;	I39495.		
DR	InterPro;	IPR003594;	Arpbind ATPase.		
DR	InterPro;	IPR002048;	EF-hand.		
DR	InterPro;	IPR003660;	HAMP.		
DR	InterPro;	IPR003661;	His_kinA.		
DR	InterPro;	IPR005467;	His_kinase.		
DR	InterPro;	IPR000014;	PAS_domain.		
DR	InterPro;	IPR002290;	Ser_thr_kinase.		
DR	Pfam;	PF00672;	HAMP; 1.		
DR	Pfam;	PF02518;	HATPase_c; 1.		
DR	Pfam;	SM00304;	HAMP; 1.		
DR	SMART;	SM00387;	HATPase_c; 1.		
DR	SMART;	SM00388;	Hiska; 1.		

P37800 sporobolomy
P52273 bombyx mori
Q9W577 gallus galli
P32262 ovis aries
P49595 caenorhabdi
P41981 synechocyst
P40688 drosophila
P49699 oryctolagus
Q13671 homo sapien
Q897U7 xanthomonas
O70507 mus musculus

34 32 69.6 322 1 ALDX SPOSA
35 32 69.6 450 1 TEA_BOMMO
36 32 69.6 461 1 SX10_CHICK
37 32 69.6 465 1 ANT3_SHEEP
38 32 69.6 491 1 P2C1_CABEL
39 32 69.6 491 1 YK5_CABEL
40 32 69.6 518 1 GGT_SINY3
41 32 69.6 548 1 SNA_DROME
42 32 69.6 709 1 ANDR_RABT
43 32 69.6 783 1 RIN1_HUMAN
44 32 69.6 916 1 IF2_XANCP
45 32 69.6 1186 1 HCN4_MOUSE

RA Kiuchi Y., Suzuki H., Hirohashi T., Tyson C.A., Sugiyama Y.;
RT "cDNA cloning and inducible expression of human multidrug resistance
EL associated protein 3 (MRP3).";
RL FEBS Lett. 433:149-152(1998).
RN (2)
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=99197095; PubMed=10094960;
RA Koenig J., Rost D., Cui Y., Keppler D.;
RT "Characterization of the human multidrug resistance protein isoform
EL MRP3 localized to the basolateral hepatocyte membrane.";
RL Hepatology 29:1156-1163(1999).
RN (3)
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=99032812; PubMed=9813153;
RA Uchiuni T., Hinoehita E., Haga S., Nakamura T., Tanaka T., Toh S.,
EL Furuhashi M., Kawabe T., Wada M., Kagotani K., Okumura K., Kohno K.,
RA Akiyama S., Kuwano M.;
RT "Isolation of a novel human canalicular multispecific organic anion
EL transporter, cMOAT2/MRP3, and its expression in cisplatin-resistant
RT cancer cells with decreased ATP-dependent drug transport.";
RL Biochem. Biophys. Res. Commun. 252:103-110(1998).
RN (4)
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Kool M., de Haas M., Ponne N.J., Baas F., Borst P.;
RT "Complete coding sequence of human MRP3, a homolog of the human
EL multidrug resistance-associated protein MRP1.";
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN (5)
RP SEQUENCE FROM N.A. (ISOFORMS 3A AND 3B).
RC TISSUE=Liver;
RX MEDLINE=99107643; PubMed=9889399;
RA Fromm M.F., Leake B., Roden D.M., Wilkinson G.R., Kim R.B.;
RT "Human MRP3 transporter: identification of the 5'-flanking region,
EL genomic organization and alternative splice variants.";
RL Biochim. Biophys. Acta 1415:369-374(1999).
RN (6)
RP SEQUENCE FROM N.A. (ISOFORM 3S1).
RC TISSUE=Colon carcinoma;
RA Auclair D., Alonso E., Chen L.B.;
RT "Identification of a novel splice variant of MRP3 involved in
EL resistance to DNA damaging agents.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN (7)
RP SEQUENCE FROM N.A.
RC MEDLINE=99043202; PubMed=9827529;
RA Belinsky M.G., Bain L.J., Balsara B.B., Testa J.R., Kruh G.D.;
RT "Characterization of MOAT-C and MOAT-D, new members of the MRP/cMOAT
EL subfamily of transporter proteins.";
RL J. Natl. Cancer Inst. 90:1735-1741(1998).
RN (8)
RP SEQUENCE OF 1043-1527 FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=97413640; PubMed=9270026;
RA Kool M., de Haas M., Scheffer G.L., Scheper R.J., van Eijk M.J.,
EL Juijn J.A., Baas F., Borst P.;
RT "Analysis of expression of cMOAT (MRP2), MRP3, MRP4, and MRP5,
EL homologues of the multidrug resistance-associated protein gene
EL (MRP1), in human cancer cell lines.";
RL Cancer Res. 57:3537-3547(1997).
CC -1- FUNCTION: MAY ACT AS AN INDUCIBLE TRANSPORTER IN THE BILIARY AND
CC INTESTINAL EXCRETION OF ORGANIC ANIONS.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=4;
CC Comment-Additional isoforms seem to exist;
CC Name=3;
CC IsoId=O15438-1; Sequence=Displayed;
CC Name=3A;
CC IsoId=O15438-2; Sequence=VSP_000042;
CC Name=3B;
CC

CC IsoId=O15438-3; Sequence=VSP_000040, VSP_000041;
CC Name=3L;
CC IsoId=O15438-4; Sequence=VSP_000043;
CC -1- TISSUE SPECIFICITY: MAINLY EXPRESSED IN THE LIVER. ALSO EXPRESSED
CC IN SMALL INTESTINE, COLON, PROSTATE, TESTIS, BRAIN AND AT A LOWER
CC LEVEL IN THE KIDNEY.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MRP SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC

DR EMBL; AB010887; BAA28146.1; -
DR EMBL; Y17511; CAA76558.2; -
DR EMBL; AF093552; AAC34668.1; -
DR EMBL; AF009670; AAD01430.1; -
DR EMBL; AF085690; AAD02845.1; -
DR EMBL; AF085691; AAD02846.1; -
DR EMBL; AF085692; AAD02847.1; -
DR EMBL; AF154001; AAD38185.1; -
DR EMBL; AF104943; AAD04170.1; -
DR EMBL; U83639; AAB71756.1; -
DR PIR; JEO336; JEO336.
DR HSSP; P13589; INED.
DR Genew; HGNC:54; ABCC3.
DR MIM; 604323; -
DR GO; GO:000587; C:integral to plasma membrane; TAS.
DR GO; GO:0005624; C:membrane fraction; TAS.
DR GO; GO:0005524; F:ATP binding activity; TAS.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. .; TAS.
DR GO; GO:0008514; F:organic anion transporter activity; TAS.
DR GO; GO:0006832; P:small molecule transport; TAS.
DR InterPro; IPR003593; AAA ATPase.
DR InterPro; IPR001140; ABC TM transp.
DR InterPro; IPR003439; ABC transporter.
DR InterPro; IPR005232; MRP_assoc.
DR Pfam; PF00664; ABC_tran; 2.
DR Pfam; PF00005; ABC_tran; 2.
DR ProDom; PD000006; ABC_tran; 2.
DR SMART; SM00382; AAA; 2.
DR TIGRFAMs; TIGR00957; MRP_assoc_pro; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 2.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 2.
KW ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat;
KW Alternative splicing.
FT DOMAIN 1 32 EXTRACELLULAR (BY SIMILARITY).
FT TRANSMEM 33 53 1 (BY SIMILARITY).
FT DOMAIN 54 73 CYTOPLASMIC (BY SIMILARITY).
FT TRANSMEM 74 94 2 (BY SIMILARITY).
FT DOMAIN 95 99 EXTRACELLULAR (BY SIMILARITY).
FT TRANSMEM 100 120 3 (BY SIMILARITY).
FT DOMAIN 121 132 CYTOPLASMIC (BY SIMILARITY).
FT TRANSMEM 133 153 4 (BY SIMILARITY).
FT DOMAIN 154 171 EXTRACELLULAR (BY SIMILARITY).
FT TRANSMEM 172 192 5 (BY SIMILARITY).
FT DOMAIN 193 302 CYTOPLASMIC (BY SIMILARITY).
FT TRANSMEM 303 323 6 (BY SIMILARITY).
FT DOMAIN 324 349 EXTRACELLULAR (BY SIMILARITY).
FT TRANSMEM 350 370 7 (BY SIMILARITY).
FT DOMAIN 371 426 CYTOPLASMIC (BY SIMILARITY).
FT TRANSMEM 427 447 8 (BY SIMILARITY).
FT DOMAIN 448 450 EXTRACELLULAR (BY SIMILARITY).
FT TRANSMEM 451 471 9 (BY SIMILARITY).
FT DOMAIN 472 533 CYTOPLASMIC (BY SIMILARITY).
FT TRANSMEM 534 554 10 (BY SIMILARITY).
FT DOMAIN 555 576 EXTRACELLULAR (BY SIMILARITY).
FT TRANSMEM 577 597 11 (BY SIMILARITY).
FT DOMAIN 598 963 CYTOPLASMIC (BY SIMILARITY).
FT TRANSMEM 964 984 12 (BY SIMILARITY).

"The allele-specific synthetic lethality of *p1A-p1g3* double mutants predicts interactive domains of SecY and SecE";
EMBO J. 14:884-893(1995).
-!- FUNCTION: ESSENTIAL FOR PROTEIN EXPORT.
-!- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLLOCATION APPARATUS WHICH COMPRISE SECA, SECB, SECD, SECE, SECF, SECG AND SECH. THE TRANSLLOCATION CHANNEL SEEMS TO BE COMPOSED OF A SECA HOMODIMER AND FOUR COPIES OF A SECEG COMPLEX.
-!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
-!- SIMILARITY: Belongs to the secE/SEC61-gamma family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; M30610; AAC24621.1; -;
EMBL; U00006; AAC43079.1; -;
EMBL; AE000472; AAC76955.1; -;
EMBL; AE016770; AAN83364.1; -;
EMBL; AE005629; AAG59177.1; -;
EMBL; AP002567; BAB38327.1; -;
PIR; A35139; VXCSE;
PIR; E86089; E86089;
PIR; H91241; H91241;
ECOGene; EG10939; secE.
InterPro; IPR001901; SecE.
InterPro; IPR005807; SecE_bac.
PIR; PF00584; SecE; 1.
PRINTS; PR01650; SECETRNLCASE.
TIGRFAMS; TIGR00964; 3a0501s06; 1.
PROSITE; PS01067; SECE_SECG1G; 1.
Protein transport; Translocation; Transmembrane; Inner membrane;
Complete proteome.
DOMAIN 1 18 CYTOPLASMIC (PROBABLE).
DOMAIN 19 36 PROBABLE.
DOMAIN 37 44 PERIPLASMIC (PROBABLE).
DOMAIN 45 63 PROBABLE.
DOMAIN 64 92 CYTOPLASMIC (PROBABLE).
DOMAIN 93 111 PROBABLE.
DOMAIN 112 127 PERIPLASMIC (PROBABLE).
SEQUENCE 127 AA; 13643 MW; 94D37280522875CE CRC64;

Query Match 73.9%; Score 34; DB 1; Length 127;
Best Local Similarity 66.7%; Pred. No. 16;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 STEGEGSGR 9
Db 4 NTEAQSGSR 12

RESULT 5
SECE_SALTY STANDARD; PRT; 127 AA.
AC Q919K1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Preprotein translocase secE subunit
GN SECE OR STM4147 OR STMFL6 OR STY3738 OR T3480.
OS Salmonella typhimurium, and
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
ON NCBI_TaxID=602, 601;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;

RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P., Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D., Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E., Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M., Waterston R., Wilson R.K.;
RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2."; Nature 413:852-856(2001).
RL Nature 413:852-856(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J., Churcher C., Bentley K.D., Bentley S.D., Holden M.T.G., Sebaihia M., Baker S., Basham D., Brooks K., Chillingworth T., Connor P., Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J., Feltwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K., Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C., Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.;
RT "Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18."; Nature 413:848-852(2001).
RL Nature 413:848-852(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=TY2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J., Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of *Salmonella enterica* serovar Typhi strains Ty2 and CT18."; J. Bacteriol. 185:2330-2337(2003).
RL J. Bacteriol. 185:2330-2337(2003).
CC -!- FUNCTION: ESSENTIAL FOR PROTEIN EXPORT (By similarity).
CC -!- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLLOCATION APPARATUS WHICH COMPRISE SECA, SECB, SECD, SECE, SECF, SECG AND SECH. THE TRANSLLOCATION CHANNEL SEEMS TO BE COMPOSED OF A SECA HOMODIMER AND FOUR COPIES OF A SECEG COMPLEX (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane (By similarity).
CC -!- SIMILARITY: Belongs to the secE/SEC61-gamma family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AF170176; AAF33494.1; -;
EMBL; AE008893; AAL22975.1; -;
EMBL; AL627279; CAD09493.1; -;
EMBL; AB016845; AAO70996.1; -;
StyGene; SG????; secE.
InterPro; IPR001901; SecE.
PIR; PF00584; SecE; 1.
PRINTS; PR01650; SECETRNLCASE.
TIGRFAMS; TIGR00964; 3a0501s06; 1.
PROSITE; PS01067; SECE_SECG1G; 1.
Protein transport; Translocation; Transmembrane; Inner membrane;
Complete proteome.
DOMAIN 1 18 CYTOPLASMIC (PROBABLE).
DOMAIN 19 36 PROBABLE.
DOMAIN 37 44 PERIPLASMIC (PROBABLE).
DOMAIN 45 63 PROBABLE.
DOMAIN 64 92 CYTOPLASMIC (PROBABLE).
DOMAIN 93 111 PROBABLE.
DOMAIN 112 127 PERIPLASMIC (PROBABLE).
SEQUENCE 127 AA; 13689 MW; 640DB5C2080E775D CRC64;

Query Match 73.9%; Score 34; DB 1; Length 127;
Best Local Similarity 66.7%; Pred. No. 16;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 STEGEGSGR 9
Db 4 NTEAQSGSR 12

RESULT 5
SECE_SALTY STANDARD; PRT; 127 AA.
AC Q919K1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Preprotein translocase secE subunit
GN SECE OR STM4147 OR STMFL6 OR STY3738 OR T3480.
OS Salmonella typhimurium, and
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
ON NCBI_TaxID=602, 601;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;

AC P52178;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Triose phosphate/phosphate translocator, non-green plastid,
 GN Chloroplast precursor (CTPT).
 GN NGTPT.
 OS Brassica oleracea (Cauliflower).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosidsII; Brassicales; Brassicaceae; Brassica.
 OX NCBI_taxid=3712;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Bettytis; TISSUE=Flower;
 RX MEDLINE=9724325; PubMed=9090886;
 RA Fischer K., Kammerer B., Gutensohn M., Arbinger B., Weber A.,
 RA Hausler R.E., Fluegge U.-I.,
 RT "A new class of plastidic phosphate translocators: a putative link
 RT between primary and secondary metabolism by the
 RT phenoenolpyruvate/phosphate antiporter.";
 RL Plant Cell 9:453-462(1997).
 CC -!- FUNCTION: MEDIATES THE EXPORT OF FIXED CARBONS FROM THE
 CC CHLOROPLASTS INTO THE CYTOSOL IN THE FORM OF TRIOSE PHOSPHATES.
 CC -!- SUBUNIT: Homodimer (By similarity).
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. LOCATED IN
 CC ZONES OF CONTACT BETWEEN THE INNER AND OUTER MEMBRANE OF THE
 CC CHLOROPLAST (BY SIMILARITY).
 CC -!- SIMILARITY: STRONG, TO OTHER PLANTS CTPT. ALSO SIMILAR TO
 CC YEAST PROTEIN SLV41.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; UL3632; AAA84892.1; .
 DR PIR; T14438; T14438.
 DR InterPro; IPR000820; DUF6.
 DR Pfam; PF00892; DUF6; 1.
 DR TIGRFAMs; TIGR00817; tpt; 1.
 KW Transmembrane, Chloroplast; Transit peptide; Transport.
 FT TRANSIT 1 82
 FT CHAIN 83 402
 FT
 FT DOMAIN 83 99
 FT TRANSMEM 99 119
 FT DOMAIN 120 131
 FT TRANSMEM 132 152
 FT DOMAIN 153 209
 FT TRANSMEM 210 230
 FT DOMAIN 231 278
 FT TRANSMEM 279 298
 FT DOMAIN 299 374
 FT TRANSMEM 375 394
 FT DOMAIN 395 402
 SQ SEQUENCE 402 AA; 43523 MW; 4970CCFF6E77CB03 CRC64;
 Query Match 73.9%; Score 34; DB 1; Length 402;
 Best Local Similarity 85.7%; Pred. No. 53;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 EGEGR 9
 |||||:
 Db 89 EGEGR 95

RESULT 9
 KDC2_DROME

ID KDC2_DROME STANDARD; PRT; 502 AA.
 AC P16912;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Protein kinase DC2 (EC 2.7.1.1-).
 GN PKA-C3 OR DC2.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_taxid=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Canton-S;
 RX MEDLINE=89107390; PubMed=3215511;
 RA Kalderson D., Rubin G.M.;
 RT "Isolation and characterization of Drosophila cAMP-dependent protein
 RT kinase genes.";
 RL Genes Dev. 2:1539-1556(1988).
 CC -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
 CC -!- TISSUE SPECIFICITY: MORE ABUNDANT IN ADULT HEAD THAN ADULT BODY.
 CC -!- DEVELOPMENTAL STAGE: IN EMBRYONS, PUPAE AND ADULTS.
 CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC CAMP SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; X16961; CAA34835.1; .
 DR HSP; P05132; IATP.
 DR FlyBase; FBgn000489; Pka-C3.
 DR GO; GO:0008602; P: cAMP-dependent protein kinase, catalyst act. .; IDA.
 DR GO; GO:0006468; P: protein amino acid phosphorylation; IDA.
 DR InterPro; IPR000961; Pkinase C.
 DR InterPro; IPR000719; Prot_kinase.
 DR InterPro; IPR002290; Ser_thr_kinase.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF00433; pkinase C; 1.
 DR PRINTS; PR00109; TYRKINASE.
 DR ProDom; PD000001; Prot_kinase; 1.
 DR SMART; SM00133; S_TK_X; 1.
 DR SMART; SM00220; S_TKC; 1.
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE; PS00108; PROTEIN KINASE ST; 1.
 DR PROSITE; PS00111; PROTEIN KINASE DOM; 1.
 KW Transferase; Serine/threonine-protein kinase; ATP-binding.
 FT DOMAIN 193 447
 FT NP_BIND 193 207
 FT BINDING 222 222
 FT ACT_SITE 316 316
 SQ SEQUENCE 502 AA; 56960 MW; CAAC474A5A40193D CRC64;
 Query Match 73.9%; Score 34; DB 1; Length 502;
 Best Local Similarity 85.7%; Pred. No. 67;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 EGEGR 9
 |||||:
 Db 137 EGEGR 143

RESULT 10
 CYAL_DROME

ID CYAL_DROME STANDARD; PRT; 2248 AA.
 AC P32870;
 DT 01-OCT-1993 (Rel. 27, Created)

```

RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RA "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
CC -!- CATALYTIC ACTIVITY: Hydrolyzes proteins with a preference for Tyr
CC or Phe in the P1' position. Has no action on amino-acid p-
CC nitroanilides.
CC -!- COFACTOR: BINDS 1 ZINC ION (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- PTM: THE N-TERMINUS IS BLOCKED.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M7.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M81703; AAA26740.1; -
CC EMBL; M89476; AAA26805.1; -
CC EMBL; D00670; BAA00573.1; -
CC EMBL; AL939131; CAB76351.1; -
CC PIR; JH0571; JH0571.
CC HSP; P56406; 1KHJ.
CC MEROPS; M07.001; .
CC InterPro; IPR006026; Nzn_Mtpeptdse.
CC InterPro; IPR000013; Peptidase_M7.
CC InterPro; IPR006025; Zn_Mtpeptdse.
CC Pfam; PF02031; Peptidase_M7; 1.
CC PRINTS; PR00787; NEUTRALPTASE.
CC PRODOM; PD016028; Peptidase_M7; 1.
CC SMART; SM00235; ZnMG_1.
CC PROSITE; PS00142; ZINC_PROTEASE; FALSE NEG.
CC KW Hydrolyase; Metalloprotease; Zinc; Signal; Zymogen; Complete proteome.
FT SIGNAL 1 29
FT PROPEP 30 42
FT CHAIN 43 227
FT METAL 163 163
FT ACT_SITE 164 164
FT METAL 167 167
FT METAL 173 173
FT DISULFID 179 192
FT CONFLICT 210 210
FT SEQUENCE 227 AA; 23765 MW; 7CAE68EADD8678CF CRC64;
Query Match 71.7%; Score 33; DB 1; Length 227;
Best Local Similarity 75.0%; Pred. No. 44;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 STEGEGSG 8
DB 136 STDGHGSG 143
RESULT 12
VALI_CSMV STANDARD; PRT; 295 AA.
AC P18921;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DE All protein (33.2 kDa protein).
OS Chloris striata mosaic virus (CSMV).
OC Viruses; ssDNA viruses; Geminiviridae; Mastrevirus.
OX NCBI_TaxID=10820;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88219532; PubMed=3369088;
RA Andersen M.T., Richardson K.A., Harbison S.A., Morris B.A.M.;
RT "Nucleotide sequence of the geminivirus chloris striata mosaic

```

```

RT virus.";
RL Virology 164:443-449(1988).
CC -!- SIMILARITY: BELONGS TO GEMINIVIRUSES AL1 PROTEIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M20021; -; NOT_ANNOTATED_CDS.
CC PIR; JU0043; JU0043; Gemini_A11.
CC InterPro; IPR001191; Gemini_A11.
CC Pfam; PF00799; Gemini_A11; 1.
CC PRINTS; PR00273; GEMCOATALL.
CC PRODOM; PD000736; Gemini_A11; 1.
CC SEQUENCE 295 AA; 33156 MW; 3386AD9080610B34 CRC64;
Query Match 71.7%; Score 33; DB 1; Length 295;
Best Local Similarity 85.7%; Pred. No. 59;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 TREGSG 8
DB 9 SEGEGSG 15
RESULT 13
FMT BRUME STANDARD; PRT; 306 AA.
ID FMT BRUME STANDARD; PRT; 306 AA.
AC Q8YDB3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Methionyl-tRNA formyltransferase (EC 2.1.2.9).
GN FMT OR BMEI10265 OR BRA1034.
OS Brucella melitensis, and
OS Brucella suis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29459, 29461;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=B.melitensis; STRAIN=16M / ATCC 23456 / Biotype 1;
RX MEDLINE=20020109; PubMed=11756688;
RA DelVecchio V.G., Kapatral V., Redkar R.J., Patra G., Mujar C., Los T.,
RA Ivanova N., Anderson I., Bhattacharyya A., Lykidis A., Reznik G.,
RA Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltsman E.,
RA Seikov E., Elzer F.H., Ragius S., O'Callaghan D., Letesson J.-J.,
RA Haselkorn R., Kyrpides N., Overbeek R.;
RT "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=B.suis; STRAIN=1330 / Biovar 1;
RX MEDLINE=22247741; PubMed=12271122;
RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
RA Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beanan M.J.,
RA Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,
RA Nelson W.C., Ayodeji B., Kraul M., Shetty J., Malek J., Van Aken S.E.,
RA Riedmuller S., Tettelin H., Gill S.R., White O., Salzberg S.L.,
RA Hoover D.L., Lindler L.E., Halling S.M., Boyle S.M., Fraser C.M.;
RT "The Brucella suis genome reveals fundamental similarities between
RT animal and plant pathogens and symbionts.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).
CC -!- FUNCTION: Modify the free amino group of the aminoacyl moiety of
CC methionyl-tRNA(fMet). The formyl group appears to play a dual role
CC in the initiator identity of N-formylmethionyl-tRNA by (I)
CC promoting its recognition by IF2 and (II) impairing its binding to
CC EFTu-GTP (By similarity).

```

CC -!- SUBCELLULAR LOCATION: SYNAPTIC VESICLES AND CHROMAFFIN GRANULES.
CC -!- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN PHYLOGENETICALLY
CC OLDER BRAIN REGIONS SUCH AS THE SPINAL CORD, BRAIN STEM AND
CC CEREBELLUM.
CC -!- DOMAIN: The first C2 domain mediates Ca(2+)-dependent phospholipid
CC binding.
CC -!- DOMAIN: The second C2 domain mediates interaction with Stonin 2
CC (By similarity).
CC -!- SIMILARITY: Contains 2 C2 domains.
CC -!- SIMILARITY: BELONGS TO THE SYNAPTOTAGMIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M64488; AAA63502.1; -.
CC PIR; A39454; BMT2Y.
CC HSP; P21707; 1BYN.
CC InterPro; IPR000008; C2.
CC InterPro; IPR002149; LRI.
CC InterPro; IPR001565; Synaptotagmin.
CC Pfam; PF00168; C2; 2.
CC PRINTS; PR00360; C2DOMAIN.
CC PRINTS; PR00339; SYNAPTOTAGMN.
CC SMART; SM00239; C2; 2.
CC PROSITE; PS00499; C2 DOMAIN 1; 2.
CC PROSITE; PS00004; C2 DOMAIN 2; 2.
CC Transmembrane, Repeat; Synapse; Glycoprotein.
CC DOMAIN 1 60 VESICULAR (POTENTIAL).
CC TRANSMEM 61 87 POTENTIAL.
CC DOMAIN 88 422 CYTOPLASMIC (POTENTIAL).
CC DOMAIN 136 382 PHOSPHOLIPID BINDING (PROBABLE).
CC DOMAIN 156 245 C2 DOMAIN 1.
CC DOMAIN 257 378 C2 DOMAIN 2.
CC CARBOHYD 32 32 N-LINKED (GLCNAC...) (POTENTIAL).
CC SEQUENCE 422 AA; 47209 MW; D852AF5387E0C7FD CRC64;

CC Query Match 71.7%; Score 33; DB 1; Length 422;
CC Best Local Similarity 85.7%; Pred. No. 86;
CC Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

CC QY 2 TEGEGSG 8
CC DB 128 TEGEGEG 134

Search completed: December 11, 2003, 18:20:59
Job time : 7.66667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 18:03:30 ; Search time 20.333 Seconds
(without alignments)
114.220 Million cell updates/sec

Title: US-10-014-658-6

Perfect score: 46

Sequence: 1 STEGEGSGR 9

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 23:**

- 1: sp_archaea:**
- 2: sp_bacteria:**
- 3: sp_fungi:**
- 4: sp_human:**
- 5: sp_invertebrate:**
- 6: sp_mammal:**
- 7: sp_muc:**
- 8: sp_organelle:**
- 9: sp_phase:**
- 10: sp_plant:**
- 11: sp_rodent:**
- 12: sp_virus:**
- 13: sp_vertebrate:**
- 14: sp_unclassified:**
- 15: sp_rvirus:**
- 16: sp_bacteriap:**
- 17: sp_archaeap:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	89.1	355	11 Q8R3R6	Q8r3r6 mus musculus
2	41	89.1	607	4 O75043	O75043 homo sapien
3	41	89.1	653	4 Q8N3F1	Q8n3f1 homo sapien
4	41	89.1	717	11 Q8BV73	Q8bv73 mus musculus
5	41	89.1	766	11 Q8BXK2	Q8bxk2 mus musculus
6	39	84.8	167	10 Q8SMX8	Q8smx8 arabidopsis
7	39	84.8	174	10 Q87A9	Q87a9 arabidopsis
8	39	84.8	329	11 Q8CS94	Q8cs94 mus musculus
9	39	84.8	450	11 Q8C7E0	Q8c7e0 mus musculus
10	39	84.8	453	11 Q8V150	Q8v150 mus musculus
11	39	84.8	455	11 Q8V149	Q8v149 mus musculus
12	39	84.8	712	11 Q8CDV1	Q8cdv1 mus musculus
13	39	84.8	750	11 Q8BQJ7	Q8bqj7 mus musculus
14	39	84.8	887	11 Q8CDQ0	Q8cdq0 mus musculus
15	39	84.8	907	11 Q8VBW5	Q8vbw5 mus musculus
16	38	82.6	1084	10 Q43010	Q43010 oryza sativ

17	38	82.6	1100	10 Q8S064	Q8s064 oryza sativ
18	37	80.4	370	17 Q8TWA1	Q8twa1 methanopyru
19	37	80.4	212	16 Q9PB98	Q9pb98 xyliella fas
20	36	78.3	361	16 Q8XPM6	Q8xpm6 ralstonia s
21	36	78.3	697	16 Q69854	Q69854 streptomyce
22	36	78.3	1026	5 Q9VSB9	Q9vbs9 drosophila
23	36	78.3	1062	5 Q9NHD3	Q9nhd3 drosophila
24	35	76.1	58	4 Q9Y623	Q9y623 homo sapien
25	35	76.1	204	4 Q9EH02	Q9eh02 homo sapien
26	35	76.1	283	16 Q8D4S3	Q8d4s3 vibrio vuln
27	35	76.1	321	10 Q8L569	Q8l569 oryza sativ
28	35	76.1	364	6 Q9N0Z3	Q9n0z3 bos taurus
29	35	76.1	611	2 Q9REA6	Q9rea6 enterococcu
30	35	76.1	946	10 Q22015	Q22015 cylindrothe
31	35	76.1	1514	4 Q8FQA9	Q8fqa9 homo sapien
32	34	73.9	128	5 Q8IQ40	Q8iq40 drosophila
33	34	73.9	167	10 Q94D37	Q94d37 oryza sativ
34	34	73.9	203	16 Q8UK51	Q8uks1 agrobacteri
35	34	73.9	246	16 Q9L7T4	Q9l7t4 pseudomonas
36	34	73.9	257	16 Q8ET34	Q8et34 oceanobacil
37	34	73.9	295	4 Q9H746	Q9h746 homo sapien
38	34	73.9	305	10 Q9AVU3	Q9av3 nicotiana t
39	34	73.9	305	16 Q8YU20	Q8yu20 anabaena sp
40	34	73.9	323	11 Q8CLU9	Q8cll9 mus musculu
41	34	73.9	341	5 Q95UR7	Q95ur7 schistocerc
42	34	73.9	359	2 Q8KSB5	Q8ksb5 actinobacil
43	34	73.9	361	3 Q96VC1	Q96vc1 kluveromyc
44	34	73.9	368	10 Q94CV4	Q94cv4 oryza sativ
45	34	73.9	368	16 Q8Y164	Q8y164 ralstonia s

ALIGNMENTS

RESULT 1

Q8R3R6 PRELIMINARY; PRT; 355 AA.

AC Q8R3R6;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Similar to KIAA0455 gene product (Fragment).
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
EL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC024711; AH24711.1;
FT NON_TER
SQ SEQUENCE 355 AA; 39056 MW; 88842D45D222AB0 CRC64;

Query Match 89.1%; Score 41; DB 11; Length 355;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 STEGEGSG 8
Db 230 STEGEGSG 237

RESULT 2

O75043 PRELIMINARY; PRT; 607 AA.

ID O75043
AC O75043;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Hypothetical protein KIAA0455 (Fragment).
GN KIAA0455;
OS Homo sapiens (Human).

CC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
CX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Landsberg erecta; TISSUE=Flower;
RX MEDLINE=97446501; PubMed=9301089;
RA Cardon G.H., Hoehmann S., Nettlesheim K., Saedler H., Huijser P.;
RT "Functional analysis of the Arabidopsis thaliana SBP-box gene SPL3: a novel gene involved in the floral transition.";
RL Plant J. 12:367-377(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Landsberg erecta; TISSUE=Flower;
RX MEDLINE=99453765; PubMed=10524240;
RA Cardon G.H., Hoehmann S., Klein J., Nettlesheim K., Saedler H., Huijser P.;
RT "Molecular characterization of the Arabidopsis SBP-box genes.";
RL Gene 237:91-104(1999).
DR EMBL; AJ011632; CAB56584.1; -;
DR InterPro; IPR004333; SBP_plant_prot.
DR Pfam; PF03110; SBP; 1.
FT NON TER 1
SQ SEQUENCE 167 AA; 19303 MW; 1BE4F4EB258EF742 CRC64;
Query Match 84.8%; Score 39; DB 10; Length 167;
Best Local Similarity 88.9%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 STEGEGSGR 9
DB 125 STYEGSGR 133
RESULT 7
Q9S7A9 PRELIMINARY; PRT; 174 AA.
AC Q9S7A9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Squamosa promoter binding protein-like 4 (F12M16.2) (Transcription factor, putative).
GN SPL4.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
CX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia; TISSUE=Flower;
RX MEDLINE=97446501; PubMed=9301089;
RA Cardon G.H., Hoehmann S., Nettlesheim K., Saedler H., Huijser P.;
RT "Functional analysis of the Arabidopsis thaliana SBP-box gene SPL3: a novel gene involved in the floral transition.";
RL Plant J. 12:367-377(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia; TISSUE=Flower;
RX MEDLINE=99453765; PubMed=10524240;
RA Cardon G.H., Hoehmann S., Klein J., Nettlesheim K., Saedler H., Huijser P.;
RT "Molecular characterization of the Arabidopsis SBP-box genes.";
RL Gene 237:91-104(1999).
RN [3]
RP SEQUENCE FROM N.A.
RA Kim C., Brooks S., Buehler E., Chao Q., Dunn P., Khan S., Shinn P., Altafi H., Araujo R., Conn L., Conway A.P., Gonzalez A., Hansen N.F., Ruiz L., Kremenetskaia I., Ilenz C., Li J., Liu S., Lueros S., Rowley D., Schwartz J., Toriumi M., Vysotskaia V., Yu G., Davis R.W., Federspiel N.A., Theologis A., Ecker J.R.;
RT "Genomic sequence for Arabidopsis thaliana BAC F12M16 from chromosome 1.";
RN [4]
RP Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Haas B.J., Volfovsky N., Town C.D., Troupkan M., Alexandrov N., Alexandrov N., Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;
RT "Full-length messenger RNA sequences greatly improve genome annotation.";
RN [5]
RP Genome Biol. 0:0-0(2002).
RN [5]
RP SEQUENCE FROM N.A.
RA Brover V., Troupkan M., Alexandrov N., Lu Y.-P., Flavell R., Feldmann K.;
RT "Full-length cDNA from Arabidopsis thaliana.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ011631; CAB56583.1; -;
DR EMBL; AJ011630; CAB56582.1; -;
DR EMBL; AC008007; RAF9527.1; -;
DR EMBL; AY084902; AAM61465.1; -;
DR InterPro; IPR004333; SBP_plant_prot.
DR Pfam; PF03110; SBP; 1.
SQ SEQUENCE 174 AA; 20119 MW; B569272D389655AB CRC64;
Query Match 84.8%; Score 39; DB 10; Length 174;
Best Local Similarity 88.9%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 STEGEGSGR 9
DB 132 STYEGSGR 140
RESULT 8
Q9CS94 PRELIMINARY; PRT; 329 AA.
AC Q9CS94;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE S730403013rik protein (fragment).
GN S730403013rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryo;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S., Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I., Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R., Kadori K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H., Kuehl P., Lewis S., Matsuo Y., Nikaido I., Resole G., Quackenbush J., Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T., Blake K., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F., Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M., Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P., Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F., Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L., Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S., Hayaishizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK017487; BAB30768.1; -;
DR HSSP; P07155; IHMF.
DR MGD; MGI:191758; 5730403013rik.
DR InterPro; IPR000910; HMG_12_box.
DR Pfam; PF00505; HMG_box; 1.

CC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
CX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Landsberg erecta; TISSUE=Flower;
RX MEDLINE=97446501; PubMed=9301089;
RA Cardon G.H., Hoehmann S., Nettlesheim K., Saedler H., Huijser P.;
RT "Functional analysis of the Arabidopsis thaliana SBP-box gene SPL3: a novel gene involved in the floral transition.";
RL Plant J. 12:367-377(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Landsberg erecta; TISSUE=Flower;
RX MEDLINE=99453765; PubMed=10524240;
RA Cardon G.H., Hoehmann S., Klein J., Nettlesheim K., Saedler H., Huijser P.;
RT "Molecular characterization of the Arabidopsis SBP-box genes.";
RL Gene 237:91-104(1999).
DR EMBL; AJ011632; CAB56584.1; -;
DR InterPro; IPR004333; SBP_plant_prot.
DR Pfam; PF03110; SBP; 1.
FT NON TER 1
SQ SEQUENCE 167 AA; 19303 MW; 1BE4F4EB258EF742 CRC64;
Query Match 84.8%; Score 39; DB 10; Length 167;
Best Local Similarity 88.9%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 STEGEGSGR 9
DB 125 STYEGSGR 133
RESULT 7
Q9S7A9 PRELIMINARY; PRT; 174 AA.
AC Q9S7A9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Squamosa promoter binding protein-like 4 (F12M16.2) (Transcription factor, putative).
GN SPL4.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
CX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia; TISSUE=Flower;
RX MEDLINE=97446501; PubMed=9301089;
RA Cardon G.H., Hoehmann S., Nettlesheim K., Saedler H., Huijser P.;
RT "Functional analysis of the Arabidopsis thaliana SBP-box gene SPL3: a novel gene involved in the floral transition.";
RL Plant J. 12:367-377(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia; TISSUE=Flower;
RX MEDLINE=99453765; PubMed=10524240;
RA Cardon G.H., Hoehmann S., Klein J., Nettlesheim K., Saedler H., Huijser P.;
RT "Molecular characterization of the Arabidopsis SBP-box genes.";
RL Gene 237:91-104(1999).
RN [3]
RP SEQUENCE FROM N.A.
RA Kim C., Brooks S., Buehler E., Chao Q., Dunn P., Khan S., Shinn P., Altafi H., Araujo R., Conn L., Conway A.P., Gonzalez A., Hansen N.F., Ruiz L., Kremenetskaia I., Ilenz C., Li J., Liu S., Lueros S., Rowley D., Schwartz J., Toriumi M., Vysotskaia V., Yu G., Davis R.W., Federspiel N.A., Theologis A., Ecker J.R.;
RT "Genomic sequence for Arabidopsis thaliana BAC F12M16 from chromosome 1.";
RN [4]
RP Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Haas B.J., Volfovsky N., Town C.D., Troupkan M., Alexandrov N., Alexandrov N., Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;
RT "Full-length messenger RNA sequences greatly improve genome annotation.";
RN [5]
RP Genome Biol. 0:0-0(2002).
RN [5]
RP SEQUENCE FROM N.A.
RA Brover V., Troupkan M., Alexandrov N., Lu Y.-P., Flavell R., Feldmann K.;
RT "Full-length cDNA from Arabidopsis thaliana.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ011631; CAB56583.1; -;
DR EMBL; AJ011630; CAB56582.1; -;
DR EMBL; AC008007; RAF9527.1; -;
DR EMBL; AY084902; AAM61465.1; -;
DR InterPro; IPR004333; SBP_plant_prot.
DR Pfam; PF03110; SBP; 1.
SQ SEQUENCE 174 AA; 20119 MW; B569272D389655AB CRC64;
Query Match 84.8%; Score 39; DB 10; Length 174;
Best Local Similarity 88.9%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 STEGEGSGR 9
DB 132 STYEGSGR 140
RESULT 8
Q9CS94 PRELIMINARY; PRT; 329 AA.
AC Q9CS94;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE S730403013rik protein (fragment).
GN S730403013rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryo;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S., Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I., Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R., Kadori K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H., Kuehl P., Lewis S., Matsuo Y., Nikaido I., Resole G., Quackenbush J., Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T., Blake K., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F., Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M., Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P., Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F., Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L., Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S., Hayaishizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK017487; BAB30768.1; -;
DR HSSP; P07155; IHMF.
DR MGD; MGI:191758; 5730403013rik.
DR InterPro; IPR000910; HMG_12_box.
DR Pfam; PF00505; HMG_box; 1.

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 STEGEGSGR 9
| | | | | | | |
DB 11 STEGEGDGK 19

RESULT 13

Q8BQJ7 PRELIMINARY; PRT; 750 AA.
AC Q8BQJ7; 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE Bobby Sox homolog (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CS7BL/6J; TISSUE=Body;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
DR EMBL; AK049516; BAC33788.1; --
FT NON TER 750 750
SQ SEQUENCE 750 AA; 83985 MW; 5035FF87DBAFAF10 CRC64;

Query Match 84.8%; Score 39; DB 11; Length 750;

Best Local Similarity 77.8%; Pred. No. 65;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 STEGEGSGR 9
| | | | | | | |
DB 11 STEGEGDGK 19

RESULT 14

Q8CDQ0 PRELIMINARY; PRT; 887 AA.
AC Q8CDQ0; 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE Bobby Sox homolog.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CS7BL/6J; TISSUE=Testis;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
DR EMBL; AK029747; BAC26596.1; --
SQ SEQUENCE 887 AA; 98500 MW; B1913D47BAE14B1C CRC64;

Query Match 84.8%; Score 39; DB 11; Length 887;

Best Local Similarity 77.8%; Pred. No. 78;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 STEGEGSGR 9
| | | | | | | |
DB 11 STEGEGDGK 19

RESULT 15

Q8VW5 PRELIMINARY; PRT; 907 AA.
AC Q8VW5; 01-MAR-2002 (TREMELrel. 20, Created)
DT 01-MAR-2002 (TREMELrel. 20, Last sequence update)
DT 01-OCT-2002 (TREMELrel. 22, Last annotation update)
DE HMG-box transcription factor BBXa.
GN 5730403O13RIK OR BBX.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Lee C.-J., Chan W.-I., Appleby V.J., Orme A.T., Scotting P.J.;
RT "BBX is expressed in developing CNS and in neuronal tumours."
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF454944; AAL58873.1; --
DR EMBL; AF454943; AAL58872.1; --
DR MGD; MGI:1917758; 5730403O13RIK.
DR InterPro; IPR000910; HMG_12_box.
DR Pfam; PF00505; HMG_box; 1.
DR SMART; SM00398; HMG; 1.
SQ SEQUENCE 907 AA; 100803 MW; 62F0EE51266176EF CRC64;

Query Match 84.8%; Score 39; DB 11; Length 907;

Best Local Similarity 77.8%; Pred. No. 80;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 STEGEGSGR 9
| | | | | | | |
DB 11 STEGEGDGK 19

Search completed: December 11, 2003, 18:28:19
Job time : 39.3333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 17:22:25 ; Search time 25 Seconds
(without alignments)
57.142 Million cell updates/sec

Title: US-10-014-658-8

Perfect score: 46

Sequence: 1 SEBEGSGR 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq 19Jun03.*

```
1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	46	100.0	9	21 AAY44472	Human antithrombin
2	40	87.0	9	21 AAY44470	Human antithrombin
3	40	87.0	9	21 AAY44471	Human antithrombin
4	39	84.8	402	18 AAW13497	Plant plasid phos
5	38	82.6	75	23 ABP32532	Human ORF1505 prot
6	38	82.6	85	22 ABG54222	Human liver peptid
7	38	82.6	85	22 ABB39272	Peptide #6778 enco
8	38	82.6	85	22 ABB24101	Protein #6100 enco
9	38	82.6	85	22 AAM59936	Human brain expres

10	38	82.6	85	22 AAW72531	Human bone marrow
11	38	82.6	85	22 AAM19618	Peptide #6052 enco
12	38	82.6	85	22 AAM32775	Peptide #6812 enco
13	38	82.6	85	23 ABG42351	Human peptid enco
14	38	82.6	152	22 ABG26522	Novel human diago
15	38	82.6	212	22 AAUI5984	Human novel secret
16	38	82.6	212	24 ABUS5053	Human novel polype
17	36	78.3	112	23 ABP05715	Human ORFX protein
18	36	78.3	117	22 AAM25577	Human protein sequ
19	36	78.3	124	22 AAE03314	Human gene 2 enco
20	36	78.3	124	23 ABG54451	Human albumin fusi
21	36	78.3	178	22 AAU27668	Human protein AFP6
22	36	78.3	178	22 AAE03375	Human gene 2 enco
23	36	78.3	482	22 ABB69968	Drosophila melanog
24	36	78.3	839	23 AAU86145	Human PRO7168 poly
25	36	78.3	839	23 ABG34056	Human Pro peptide
26	35	76.1	64	22 ABG49548	Human liver peptid
27	35	76.1	64	22 ABB29540	Peptide #2191 enco
28	35	76.1	64	22 ABB34725	Peptide #2231 enco
29	35	76.1	64	22 ABB20137	Protein #2136 enco
30	35	76.1	64	22 AAM55522	Human brain expres
31	35	76.1	64	22 AAM67906	Human bone marrow
32	35	76.1	64	22 AAM15723	Peptide #2157 enco
33	35	76.1	64	22 AAM28228	Peptide #2265 enco
34	35	76.1	64	22 AAM03457	Peptide #2139 enco
35	35	76.1	64	23 ABG37442	Human peptid enco
36	35	76.1	93	23 ABB34334	Human synthase-lik
37	35	76.1	104	23 ABP08775	Human ORFX protein
38	35	76.1	142	21 AAG40971	Zea mays protein f
39	35	76.1	174	21 AAG07140	Arabidopsis thalia
40	35	76.1	174	21 AAG47381	Arabidopsis thalia
41	35	76.1	179	22 ABB14761	Human nervous syst
42	35	76.1	179	24 ABU70953	Human adipocyte Se
43	35	76.1	282	21 AAG07139	Arabidopsis thalia
44	35	76.1	282	21 AAG47380	Arabidopsis thalia
45	35	76.1	294	21 AAG07138	Arabidopsis thalia

ALIGNMENTS

RESULT 1

AAV44472

ID AAY44472 standard; peptide; 9 AA.

XX AAY44472;

AC AAY44472;

DT 27-MAR-2000 (first entry)

XX Human antithrombin III variant Bb.D (residues 385-393).

XX Human; antithrombin III; ATIII variant Bb.D; elastase-resistant;

XX IGG activated neutrophil resistant; anti-thrombin activity; heparin;

XX anti-factor Xa activity; blood clotting disorder; sepsis; trauma; stroke;

XX thrombin activation-related pathological symptom; restenosis; thrombolysis;

XX acute respiratory distress syndrome; thromboembolism; reocclusion.

XX Homo sapiens.

OS Synthetic.

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

XX Key

OS Homo sapiens.
 XX Synthetic.
 FH Key
 FT Location/Qualifiers
 FT Misc-difference 2 /note= "ATIII.N135A Thr at 386 is substituted by Glu"
 FT Misc-difference 3 /note= "ATIII.N135A Ala at 387 is substituted by Glu"
 FT Misc-difference 4 /note= "ATIII.N135A Val at 388 is substituted by Gly"
 FT Misc-difference 5 /note= "ATIII.N135A Val at 389 is substituted by Glu"
 FT Misc-difference 6 /note= "ATIII.N135A Ile at 390 is substituted by Ala"
 FT Misc-difference 7 /note= "ATIII.N135A Ala at 391 is substituted by Ser"
 XX WO9958098-A2.
 XX
 XX 18-NOV-1999.
 XX
 XX 12-MAY-1999; 99WO-US10549.
 XX
 XX 12-MAY-1998; 98US-0085197.
 PR 05-MAY-1999; 99US-0085197.
 XX
 XX (BOCK/) BOCK S C.
 PA (PICA/) PICARD V.
 PA (ZEND/) ZENDEHROUH P.
 XX
 XX Bock SC, Picard V, Zendeherouh P;
 XX
 XX WPI; 2000-116274/10.
 XX
 XX New modified human antithrombin III compounds, used for treating e.g.
 PT sepsis, trauma, acute respiratory distress syndrome, restenosis,
 PT thrombosis, thromboembolism or stroke -
 XX
 XX Claim 13; Page 57; 75pp; English.
 PS
 XX
 XX The present sequence is from an antithrombin III (ATIII) variant, Bb.C
 CC derived from human ATIII.N135A cDNA insert of the pBluebac baculovirus
 CC expression construct and comprises residues 385-393. The variant has
 CC improved resistance to elastase and Igg-activated neutrophils while
 CC retains anti-thrombin and anti-factor Xa activities. It may be
 CC expressed as glycoforms with enhanced heparin affinity which target the
 CC blood vessel wall more efficiently than ATIIIs with normal heparin
 CC affinity. The modified ATIIIs can be used to treat thrombin activation-
 CC related pathological symptoms due to sepsis, trauma, acute
 CC respiratory distress syndrome, restenosis, thrombosis, thromboembolism
 CC and stroke. It can also be used to reduce the risk of reocclusion
 CC and restenosis in percutaneous transluminal coronary angioplasty,
 CC thrombosis associated with surgery, ischaemia/reperfusion injury, and
 CC coagulation abnormalities in cancer or surgical patients.
 XX
 XX Sequence 9 AA;
 SQ
 Query Match 87.0%; Score 40; DB 21; Length 9;
 Best Local Similarity 88.9%; Pred. No. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 SEEGEGSGR 9
 Db 1 SEEGEGSGR 9
 RESULT 4
 AAW13497
 ID AAW13497 standard; Protein; 402 AA.
 XX
 AC AAW13497;
 XX
 DT 20-OCT-1997 (first entry)

XX Plant plastid phosphoenolpyruvate-phosphate translocator.
 DE
 XX Plastid; phosphoenolpyruvate; phosphate; PPT.
 KW
 XX Brassica oleracea.
 OS
 XX DE19600357-C1.
 PN
 XX 13-FEB-1997.
 PD
 XX 08-JAN-1996; 96DE-1000357.
 PF
 XX 08-JAN-1996; 96DE-1000357.
 PR
 XX (FLUE/) FLUEGGE U.
 PA (BADI) BASF AG.
 PA
 XX Fischer K, Fluegge U, Weber A;
 XX WPI; 1997-110255/11.
 DR N-PSDB; AAT61549.
 DR
 XX DNA encoding plant plastid phospho:enol:pyruvate-phosphate
 PT translocator - for prodn. of transgenic plants, etc.
 PT
 XX Disclosure; Page 14-16; 17pp; German.
 PS
 XX The DNA may be used for the prodn. of PPT-expressing pro-or
 CC eukaryotic cells; for expression of untranslatable RNA that
 CC inhibits the synthesis of endogenous PPT in cells by an
 CC antisense effect or by ribozyme activity; for prodn. of
 CC DNA sequences encoding a plant plastid PPT with altered
 CC specificity; for isolating DNA sequences encoding polypeptides
 CC with PPT activity; and for identifying substances that inhibit
 CC the transport of phosphoenolpyruvate through inner plastid
 CC envelope membranes.
 CC
 XX Sequence 402 AA;
 SQ
 Query Match 84.8%; Score 39; DB 18; Length 402;
 Best Local Similarity 87.5%; Pred. No. 90;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 2 BEGEGSGR 9
 Db 88 BEGEGSGK 95
 RESULT 5
 ABP32532
 ID ABP32532 standard; Protein; 75 AA.
 XX
 AC ABP32532;
 XX
 DT 08-JUL-2002 (first entry)
 DT
 XX Human ORF1505 protein, SEQ ID NO:3010.
 DE
 XX Human; ORF; open reading frame; ORFX; drug screening; diagnosis;
 KW disease monitoring; cytokine; cell proliferation; cell differentiation;
 KW immune modulation; haematopoiesis regulation; tissue growth;
 KW angiogenesis; activin; inhibitor; chemotactic; chemokinetic; haemostatic;
 KW thrombolytic; tumour inhibition; bodily characteristic; fertility;
 KW behaviour; cancer; proliferative disorder; neurological disorder;
 KW cardiovascular disease; immune system disorder; organ transplantation;
 KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;
 KW hypothyroidism; cholesterol ester storage disease; infection; vulnery;
 KW vasotropic; antipsoriatic; antidiabetic; cytostatic; neurotropic;
 KW neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;
 KW cardiant; hypotensive; antithyroid; antiinflammatory; immunomodulator;
 KW dermatological; analgesic; virucide; antibacterial; fungicide.
 XX

KW Human; foetal liver; gene expression; single exon nucleic acid probe.
XX Homo sapiens.
OS
XX WO200157277-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00669.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human fetal liver -
XX
XX Claim 27; SEQ ID NO 31907; 639pp + sequence listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX fetal liver. The present sequence is a peptide encoded by a single exon
XX nucleic acid probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 85 AA;
SQ
Query Match 82.6%; Score 38; DB 22; Length 85;
Best Local Similarity 87.5%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 2 EGEQSGR 9
DB 27 EGEQSGR 34
|||||
RESULT 8
ABB24101
ID ABB24101 standard; Protein; 85 AA.
XX
XX ABB24101;
XX
XX 23-JAN-2002 (first entry)
XX
XX Protein #6100 encoded by probe for measuring heart cell gene expression.
XX
XX Human; gene expression; heart; microarray; vascular system;
XX cardiovascular disease; hypertension; cardiac arrhythmia;
XX congenital heart disease.
XX
XX Homo sapiens.
XX
XX WO200157274-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00666.
XX
XX 04-FEB-2000; 2000US-0180312.
XX

PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488899/53.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX hearts -
XX
XX Claim 15; SEQ ID No 25871; 530pp; English.
XX
XX The present invention relates to single exon nucleic acid probes for
XX measuring human gene expression in a sample derived from human heart (see
XX ABA21535-ABA41305). The present sequence is a protein encoded by one such
XX probe. The probes may be used for predicting, measuring and displaying
XX gene expression in samples derived from the human heart via microarrays.
XX By measuring gene expression, the probes are useful for predicting,
XX diagnosing, grading, staging, monitoring and prognosing diseases of the
XX human heart and vascular system e.g. cardiovascular disease,
XX hypertension, cardiac arrhythmias and congenital heart disease.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 85 AA;
SQ
Query Match 82.6%; Score 38; DB 22; Length 85;
Best Local Similarity 87.5%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 2 EGEQSGR 9
DB 27 EGEQSGR 34
|||||
RESULT 9
AAM59936
ID AAM59936 standard; Protein; 85 AA.
XX
XX AAM59936;
XX
XX 05-NOV-2001 (first entry)
XX
XX Human brain expressed single exon probe encoded protein SEQ ID NO: 32041.
XX
XX Human; brain expressed exon; gene expression analysis; probe;
XX microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
XX epilepsy; cancer.
XX
XX Homo sapiens.
XX
XX WO200157275-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00667.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX PA

Db 27 BEGEGKGR 34
 RESULT 12
 AAM32775
 ID AAM32775 standard; Protein; 85 AA.
 XX AC
 AC AAM32775;
 DT 17-OCT-2001 (first entry)
 XX
 DE Peptide #6812 encoded by probe for measuring placental gene expression.
 XX
 KW Probe; microarray; human; placenta; antenatal diagnosis;
 KW genetic disorder.
 KW
 OS Homo sapiens.
 XX
 XX WO200157272-A2.
 PN
 XX 09-AUG-2001.
 PD
 XX 30-JAN-2001; 2001WO-US00663.
 PF
 XX 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 PA
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 PI
 XX WPI; 2001-488897/53.
 DR
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human placenta -
 PT
 XX Claim 27; SEQ ID No 33044; 654pp; English.
 PS
 XX The present invention relates to single exon nucleic acid probes (SENP;
 CC see AAI31315-AA157546). The present sequence is a peptide encoded by one
 CC such probe. The probes are useful for producing a microarray for
 CC predicting, measuring and displaying gene expression in samples derived
 CC from human placenta. The probes are useful for antenatal diagnosis of
 CC human genetic disorders.
 XX
 SQ Sequence 85 AA;
 Query Match 82.6%; Score 38; DB 22; Length 85;
 Best Local Similarity 87.5%; Fred. No. 28;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 2 BEGEGSGR 9
 Db 27 BEGEGKGR 34
 RESULT 13
 ABG42351
 ID ABG42351 standard; Peptide; 85 AA.
 XX AC
 AC ABG42351;
 DT 19-AUG-2002 (first entry)
 XX
 DE Human peptide encoded by genome-derived single exon probe SEQ ID 32016.
 XX
 KW Human; single exon probe; asthma; lung cancer; COPD; ILD;

KW chronic obstructive pulmonary disease; interstitial lung disease;
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemorrhoidosis;
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
 KW primary ciliary dyskinesia; pulmonary hypertension;
 KW hyaline membrane disease.
 XX
 OS Homo sapiens.
 XX
 XX WO200186003-A2.
 PN
 XX 15-NOV-2001.
 PD
 XX 30-JAN-2001; 2001WO-US00665.
 PF
 XX 04-FEB-2000; 2000US-180312P.
 PR 26-MAY-2000; 2000US-207456P.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-234687P.
 PR 27-SEP-2000; 2000US-236359P.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 PA
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 PI
 XX WPI; 2002-114183/15.
 DR
 XX Spatially-addressable set of single exon nucleic acid probes, used to
 PT measure gene expression in human lung samples -
 PT
 XX Claim 27; SEQ ID No 32016; 634pp; English.
 PS
 XX The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human lung comprising single exon nucleic acid probes having one of
 CC 12614 nucleic acid sequences mentioned in the specification, or their
 CC complements or the 12387 open reading frames derived from the 12614
 CC probes. Also included are a microarray comprising the novel set of
 CC probes; the novel set of probes which hybridise at high stringency to a
 CC nucleic acid expressed in the human lung; measuring gene expression in a
 CC sample derived from human lung, comprising (a) contacting the array with
 CC a collection of detectably labeled nucleic acids derived from human lung
 CC mRNA, and (b) measuring the label detectably bound to each probe of
 CC the array; identifying exons in a eukaryotic genome, comprising
 CC (a) algorithmically predicting at least one exon from genomic sequences
 CC of the eukaryote; and (b) detecting specific hybridisation of detectably
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
 CC having a fragment identical to the predicted exon, the probe is included
 CC in the above mentioned microarray, assigning exons to a single gene,
 CC comprising (a) identifying exons from genomic sequence by the method
 CC above and (b) measuring the expression of each of the exons in several
 CC tissues and/or cell types using hybridisation to a single exon
 CC microarrays having a probe with the exon, where a common pattern of
 CC expression of the exons in the tissues and/or cell types indicates that
 CC the exons should be assigned to a single gene; a peptide comprising one
 CC of 12011 sequences, mentioned in the specification, or encoded by the
 CC probes/open reading frames (ORF). The probes are used for gene
 CC expression analysis, and for identifying exons in a gene, particularly
 CC using human lung derived mRNA and for the study of lung diseases
 CC such as asthma, lung cancer, chronic obstructive pulmonary disease
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
 CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
 CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
 CC haemorrhoidosis, pulmonary histiocytosis, lymphangioleiomyomatosis,
 CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic
 CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension
 CC and hyaline membrane disease. The present sequence is a peptide/protein
 CC encoded by a single exon probe of the invention.
 CC Note: The sequence data for this patent did not form part

PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231965.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234597.
PR 25-SEP-2000; 2000US-0234598.
PR 26-SEP-2000; 2000US-0234599.
PR 27-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 29-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-02559678.
XX (HUMA-) HUMAN GENOME SCI INC.
PA Rosen CA, Barash SC, Ruben SM;
PI WPI; 2001-488783/53.
XX N-PSDB; AAS25971.
PT New nucleic acid molecules encoding 461 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives -
XX Claim 11; SEQ ID No 937; 980pp; English.
XX The invention relates to isolated nucleic acid molecules and their
XX encoded secreted proteins. The nucleic acids and proteins are used to
XX prevent, treat or ameliorate a medical condition in e.g. humans, mice,
XX rabbits, goats, horses, cats, dogs, chickens or sheep. They
XX are also used in diagnosing a pathological condition or susceptibility
XX to a pathological condition. Antibodies to the proteins can also
XX be used in alleviating symptoms associated with the disorders and in
XX diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
XX immunosorbant assays (ELISA). Disorders which are diagnosed or treated
XX include autoimmune diseases e.g. rheumatoid arthritis,
XX hyperproliferative disorders e.g. neoplasms of the breast or liver,
XX cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
XX e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
XX Alzheimer's disease, infections caused by bacteria, viruses and fungi
XX and ocular disorders e.g. corneal infection, and many other
XX disorders listed in the specification. The polypeptides can also
XX be used to aid wound healing and epithelial cell proliferation, to
XX prevent skin aging due to sunburn, to maintain organs before
XX transplantation, for supporting cell culture of primary tissues, to
XX regenerate tissues and in chemotaxis. The polypeptides can also be used
XX as a food additive or preservative to increase or decrease storage
XX capabilities, fat content, lipid, protein, carbohydrate, vitamins,
XX minerals, cofactors and other nutritional components. The present
XX sequence represents a novel secreted protein of the invention.
Query Match 82.6%; Score 38; DB 22; Length 212;
Best Local Similarity 87.5%; Pred. No. 70;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	35	76.1	143	2	US-08-942-819-14	Sequence 14, Appl
2	35	76.1	143	4	US-09-522-955A-22	Sequence 22, Appl
3	35	76.1	167	4	US-09-252-991A-17526	Sequence 17526, A
4	35	76.1	403	2	US-08-474-379C-14	Sequence 14, Appl
5	35	76.1	403	3	US-09-146-249A-14	Sequence 14, Appl
6	35	76.1	403	1	US-08-206-188B-14	Sequence 14, Appl
7	35	76.1	405	1	US-07-688-352C-14	Sequence 14, Appl
8	35	76.1	405	5	PCR-U991-02714-14	Sequence 14, Appl
9	35	76.1	493	4	US-09-252-991A-18838	Sequence 18838, A
10	35	76.1	694	4	US-09-252-991A-30587	Sequence 30587, A
11	35	76.1	984	4	US-09-287-354-2	Sequence 2, Appl
12	35	76.1	1189	4	US-09-287-354-3	Sequence 3, Appl
13	35	76.1	1189	4	US-09-287-354-4	Sequence 4, Appl
14	34	73.9	17	4	US-09-552-993-4	Sequence 4, Appl
15	34	73.9	93	2	US-08-902-623-43	Sequence 43, Appl
16	34	73.9	166	4	US-09-252-991A-20150	Sequence 20150, A
17	34	73.9	201	3	US-09-220-528-116	Sequence 116, App
18	34	73.9	246	4	US-09-252-991A-30011	Sequence 30011, A
19	34	73.9	247	4	US-09-107-532A-4408	Sequence 4408, Ap
20	34	73.9	473	1	US-08-103-739B-2	Sequence 2, Appl
21	34	73.9	473	2	US-08-474-404-2	Sequence 2, Appl
22	34	73.9	473	3	US-08-485-845-2	Sequence 2, Appl
23	34	73.9	473	2	US-08-482-714-2	Sequence 2, Appl
24	34	73.9	473	3	US-09-211-416-2	Sequence 2, Appl
25	34	73.9	473	3	US-09-059-958-2	Sequence 2, Appl
26	34	73.9	846	4	US-09-252-991A-23780	Sequence 23780, A
27	34	73.9	1898	1	US-08-056-200-94	Sequence 94, Appl

STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/09/146,249A
FILING DATE: 01-MAR-1994
CLASSIFICATION: 435
PRIORITY APPLICATION NUMBER: US 07/511,715
FILING DATE: 20-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36107
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
SEQUENCE CHARACTERISTICS:
LENGTH: 403 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-146-249A-14

Query Match 76.1%; Score 35; DB 3; Length 403;
Best Local Similarity 75.0%; Pred. No. 2.1e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SEESEGS 8
DB 219 TQEGSG 226

RESULT 6
US-08-206-188B-14
Sequence 14, Application US/08206188B
Patent No. 6100025
GENERAL INFORMATION:
APPLICANT: Wigler, Michael H.
APPLICANT: Colicelli, John J.
TITLE OF INVENTION: Cloning By Complementation and Related
TITLE OF INVENTION: Processes
NUMBER OF SEQUENCES: 84
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
FILING DATE: 01-MAR-1994
CLASSIFICATION: 435
PRIORITY APPLICATION NUMBER: US 07/511,715
FILING DATE: 20-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36107
TELEPHONE: 312/474-6300
TELEFAX: 312/474-6300

TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 403 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-206-188B-14

Query Match 76.1%; Score 35; DB 3; Length 403;
Best Local Similarity 75.0%; Pred. No. 2.1e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SEESEGS 8
DB 219 TQEGSG 226

RESULT 7
US-07-688-352C-14
Sequence 14, Application US/07688352C
Patent No. 5527896
GENERAL INFORMATION:
APPLICANT: Wigler, Michael H.
APPLICANT: Colicelli, John J.
TITLE OF INVENTION: Cloning By Complementation and Related
TITLE OF INVENTION: Processes
NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Bicknell
STREET: Two First National Plaza, 20 South Clark
STREET: Street
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/688,352C
FILING DATE: 19910419
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/511,715
FILING DATE: 20-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Borun, Michael F.
REGISTRATION NUMBER: 25447
REFERENCE/DOCKET NUMBER: 27805/30197
TELEPHONE: (312) 346-5750
TELEFAX: (312) 984-9740
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 405 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-688-352C-14

Query Match 76.1%; Score 35; DB 1; Length 405;
Best Local Similarity 75.0%; Pred. No. 2.1e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SEESEGS 8
DB 221 TQEGSG 228

```
RESULT 12
US-09-287-354-3
; Sequence 3, Application US/09287354
; Patent No. 6348348
; GENERAL INFORMATION:
; APPLICANT: THOMPSON, Catherine C.
; TITLE OF INVENTION: HUMAN HAIRLESS GENE AND PROTEIN
; FILE REFERENCE: Thompson-20263/0243435
; CURRENT APPLICATION NUMBER: US/09/287,354
; CURRENT FILING DATE: 1999-04-07
; EARLIER APPLICATION NUMBER: US 60/080,888
; EARLIER FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1189
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-287-354-3

Query Match      76.1%; Score 35; DB 4; Length 1189;
Best Local Similarity 87.5%; Pred. No. 6e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SEEGEGSG 8
Db 538 SEEGPGSG 545

RESULT 13
US-09-287-354-4
; Sequence 4, Application US/09287354
; Patent No. 6348348
; GENERAL INFORMATION:
; APPLICANT: THOMPSON, Catherine C.
; TITLE OF INVENTION: HUMAN HAIRLESS GENE AND PROTEIN
; FILE REFERENCE: Thompson-20263/0243435
; CURRENT APPLICATION NUMBER: US/09/287,354
; CURRENT FILING DATE: 1999-04-07
; EARLIER APPLICATION NUMBER: US 60/080,888
; EARLIER FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1189
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-287-354-4

Query Match      76.1%; Score 35; DB 4; Length 1189;
Best Local Similarity 87.5%; Pred. No. 6e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SEEGEGSG 8
Db 538 SEEGPGSG 545

RESULT 14
US-09-552-993-4
; Sequence 4, Application US/09552993
; Patent No. 6511967
; GENERAL INFORMATION:
; APPLICANT: Weisleder, Ralph
; APPLICANT: Basilion, James P.
; APPLICANT: Chiocca, Ennio Antonio
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR IMAGING
; FILE REFERENCE: 00786-420002
; CURRENT APPLICATION NUMBER: US/09/552,993
; CURRENT FILING DATE: 2000-04-21
; PRIOR APPLICATION NUMBER: 60/130,794
```

```
; PRIOR FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 17
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-552-993-4

Query Match      73.9%; Score 34; DB 4; Length 17;
Best Local Similarity 77.8%; Pred. No. 13;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 SEEGEGSG 9
Db 9 SEEDGEGR 17

RESULT 15
US-08-902-623-43
; Sequence 43, Application US/08902623
; Patent No. 5922545
; GENERAL INFORMATION:
; APPLICANT: MATHEAKIS, LARRY C.
; APPLICANT: DOWER, WILLIAM J.
; TITLE OF INVENTION: IN VITRO PEPTIDE AND ANTIBODY DISPLAY
; TITLE OF INVENTION: LIBRARIES
; NUMBER OF SEQUENCES: 75
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND AND TOWNSEND AND CREW LLP
; STREET: TWO EMBARCADERO CENTER, 8TH FLOOR
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: U.S.A.
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/902,623
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/586,176
; FILING DATE: 17-JAN-1996
; APPLICATION NUMBER: US 08/300,262
; FILING DATE: 02-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/144,775
; FILING DATE: US 23-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/12206
; FILING DATE: US 25-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: DUNN, TRACY J.
; REGISTRATION NUMBER: 34,587
; REFERENCE/DOCKET NUMBER: 16528X-0032300S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 326-2400
; TELEFAX: (415) 326-2422
; INFORMATION FOR SEQ ID NO: 43:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
US-08-902-623-43

Query Match      73.9%; Score 34; DB 2; Length 93;
Best Local Similarity 75.0%; Pred. No. 70;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

Result No.	Score	Query Match	Length	DB	ID	Description
1	38	82.6	85	9	US-09-864-711-3939	Sequence 3939, A
2	38	82.6	212	10	US-09-764-864-937	Sequence 937, App
3	36	78.3	839	12	US-10-237-496-54	Sequence 54, Appl
4	36	78.3	839	12	US-10-242-074-54	Sequence 54, Appl
5	36	78.3	839	12	US-10-242-505-54	Sequence 54, Appl
6	36	78.3	839	12	US-10-242-574-54	Sequence 54, Appl
7	36	78.3	839	12	US-10-243-361-54	Sequence 54, Appl
8	36	78.3	839	12	US-10-243-382-54	Sequence 54, Appl
9	36	78.3	839	12	US-10-243-402-54	Sequence 54, Appl
10	36	78.3	839	12	US-10-243-431-54	Sequence 54, Appl
11	36	78.3	839	12	US-10-245-164-54	Sequence 54, Appl
12	36	78.3	839	12	US-10-210-951-36	Sequence 36, Appl
13	36	78.3	839	12	US-10-244-972-54	Sequence 54, Appl
14	36	78.3	839	12	US-10-197-942-54	Sequence 54, Appl
15	36	78.3	839	12	US-10-211-884-36	Sequence 36, Appl

; CURRENT APPLICATION NUMBER: US/10/242,074
; CURRENT FILING DATE: 2002-09-11
; PRIOR APPLICATION NUMBER: 10/197942
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/059114
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/065027
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/086478
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
; PRIOR FILING DATE: 1998-06-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 54
; LENGTH: 839
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-242-074-54

Query Match 78.3%; Score 36; DB 12; Length 839;
Best Local Similarity 75.0%; Pred. No. 9, 1e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 BEGEGSGR 9
:||||:|
Db 817 DEGEGAGR 824

RESULT 5
US-10-242-505-54
; Sequence 54, Application US/10242505
; Publication No. US2003013898A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Eaton, Dan
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Goddard, Audrey
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Phillippe
; APPLICANT: Watanabe, Colin
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; APPLICANT: Fong, Sherman
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3630R1C25
; CURRENT APPLICATION NUMBER: US/10/242,505
; CURRENT FILING DATE: 2002-09-11
; PRIOR APPLICATION NUMBER: 10/197942
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/059114
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/065027
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/086478
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
; PRIOR FILING DATE: 1998-06-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 54
; LENGTH: 839
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-242-074-54

; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
; PRIOR FILING DATE: 1998-06-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 54
; LENGTH: 839
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-242-505-54

Query Match 78.3%; Score 36; DB 12; Length 839;
Best Local Similarity 75.0%; Pred. No. 9, 1e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 BEGEGSGR 9
:||||:|
Db 817 DEGEGAGR 824

RESULT 6
US-10-242-574-54
; Sequence 54, Application US/10242574
; Publication No. US2003013899A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Eaton, Dan
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Goddard, Audrey
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Phillippe
; APPLICANT: Watanabe, Colin
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; APPLICANT: Fong, Sherman
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3630R1C20
; CURRENT APPLICATION NUMBER: US/10/242,574
; CURRENT FILING DATE: 2002-09-11
; PRIOR APPLICATION NUMBER: 10/197942
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/059114
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/065027
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/086478
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
; PRIOR FILING DATE: 1998-06-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 54
; LENGTH: 839
; TYPE: PRT
; ORGANISM: Homo Sapien

```
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3630RIC39
; CURRENT APPLICATION NUMBER: US/10/243,402
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: 10/197942
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/059114
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/065027
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/086478
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
; PRIOR FILING DATE: 1998-06-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 54
; LENGTH: 839
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-243-402-54

Query Match      78.3%; Score 36; DB 12; Length 839;
Best Local Similarity 75.0%; Pred. No. 9,1e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      2 EREGSGR 9
      :||||:
Db      817 DEGEGR 824

RESULT 10
US-10-243-431-54
; Sequence 54, Application US/10243431
; Publication No. US20030138903A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Eaton, Dan
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Goddard, Audrey
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Phillippe
; APPLICANT: Watanabe, Colin
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; APPLICANT: Fong, Sherman
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3630RIC37
; CURRENT APPLICATION NUMBER: US/10/243,431
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: 10/197942
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/059114
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/065027
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/086478
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
; PRIOR FILING DATE: 1998-06-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 54
; LENGTH: 839
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-243-402-54

Query Match      78.3%; Score 36; DB 12; Length 839;
Best Local Similarity 75.0%; Pred. No. 9,1e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      2 EREGSGR 9
      :||||:
Db      817 DEGEGR 824

RESULT 11
US-10-245-164-54
; Sequence 54, Application US/10245164
; Publication No. US20030138904A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Eaton, Dan
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Goddard, Audrey
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Phillippe
; APPLICANT: Watanabe, Colin
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; APPLICANT: Fong, Sherman
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3630RIC76
; CURRENT APPLICATION NUMBER: US/10/245,164
; CURRENT FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: 10/197942
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/059114
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/065027
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/086478
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
; PRIOR FILING DATE: 1998-06-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 54
```

APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
APPLICANT: Fong, Sherman
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3630R1C1(US)
CURRENT FILING DATE: 2002-07-18
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059114
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063046
PRIOR FILING DATE: 1997-11-10
PRIOR APPLICATION NUMBER: 60/065027
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/079689
PRIOR FILING DATE: 1998-05-22
PRIOR APPLICATION NUMBER: 60/086478
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087607
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089801
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090557
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090689
PRIOR FILING DATE: 1998-07-01
PRIOR APPLICATION NUMBER: 60/091358
PRIOR FILING DATE: 1998-07-07
PRIOR APPLICATION NUMBER: 60/091978
PRIOR FILING DATE: 1998-08-10
PRIOR APPLICATION NUMBER: 60/099803
PRIOR FILING DATE: 1998-11-03
PRIOR APPLICATION NUMBER: 60/106932
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/115554
PRIOR FILING DATE: 1999-02-09
PRIOR APPLICATION NUMBER: 60/119342
PRIOR FILING DATE: 1999-03-12
PRIOR APPLICATION NUMBER: 60/123957
PRIOR FILING DATE: 1999-03-11
PRIOR APPLICATION NUMBER: 60/123972
PRIOR FILING DATE: 1999-04-01
PRIOR APPLICATION NUMBER: 60/131271
PRIOR FILING DATE: 1999-04-27
PRIOR APPLICATION NUMBER: 60/133459
PRIOR FILING DATE: 1999-05-11
PRIOR APPLICATION NUMBER: 60/135725
PRIOR FILING DATE: 1999-05-25
PRIOR APPLICATION NUMBER: 60/135729
PRIOR FILING DATE: 1999-05-25
PRIOR APPLICATION NUMBER: 60/135750
PRIOR FILING DATE: 1999-06-09
PRIOR APPLICATION NUMBER: 60/140653
PRIOR FILING DATE: 1999-06-22
PRIOR APPLICATION NUMBER: 60/141037
PRIOR FILING DATE: 1999-06-23
PRIOR APPLICATION NUMBER: 60/144732
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: 60/144758
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: 60/144790
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: 60/145228
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: 60/145698
PRIOR FILING DATE: 1999-07-26
PRIOR APPLICATION NUMBER: 60/146222
PRIOR FILING DATE: 1999-07-28
PRIOR APPLICATION NUMBER: 60/146843
PRIOR FILING DATE: 1999-08-03
PRIOR APPLICATION NUMBER: 60/148188
PRIOR FILING DATE: 1999-08-10
PRIOR APPLICATION NUMBER: 60/148513
PRIOR FILING DATE: 1999-08-12
PRIOR APPLICATION NUMBER: 60/149327
PRIOR FILING DATE: 1999-08-17
PRIOR APPLICATION NUMBER: 60/149395
PRIOR FILING DATE: 1999-08-17
PRIOR APPLICATION NUMBER: 60/150114
PRIOR FILING DATE: 1999-08-20
PRIOR APPLICATION NUMBER: 60/151700
PRIOR FILING DATE: 1999-08-31
PRIOR APPLICATION NUMBER: 60/151734
PRIOR FILING DATE: 1999-08-31
PRIOR APPLICATION NUMBER: 60/162506
PRIOR FILING DATE: 1999-10-29
PRIOR APPLICATION NUMBER: 60/170262
PRIOR FILING DATE: 1999-12-09
PRIOR APPLICATION NUMBER: 60/177118
PRIOR FILING DATE: 2000-01-20
PRIOR APPLICATION NUMBER: 60/179851
PRIOR FILING DATE: 2000-02-02
PRIOR APPLICATION NUMBER: 60/180921
PRIOR FILING DATE: 2000-02-08
PRIOR APPLICATION NUMBER: 60/187202
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: 60/198587
PRIOR FILING DATE: 2000-04-18
PRIOR APPLICATION NUMBER: 60/199614
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 60/206330
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/206368
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/209832
PRIOR FILING DATE: 2000-06-05
PRIOR APPLICATION NUMBER: 60/218371
PRIOR FILING DATE: 2000-07-13
PRIOR APPLICATION NUMBER: 60/222695
PRIOR FILING DATE: 2000-08-02
PRIOR APPLICATION NUMBER: 60/229896
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 60/230621
PRIOR FILING DATE: 2000-09-05
PRIOR APPLICATION NUMBER: 60/232887
PRIOR FILING DATE: 2000-09-15
PRIOR APPLICATION NUMBER: 60/235147
PRIOR FILING DATE: 2000-09-22
PRIOR APPLICATION NUMBER: 60/261878
PRIOR FILING DATE: 2001-01-12
PRIOR APPLICATION NUMBER: 60/261910
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: 60/261939
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: 60/262150
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: 60/264395
PRIOR FILING DATE: 2001-01-25
PRIOR APPLICATION NUMBER: 60/266421
PRIOR FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: 60/267623
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/274399
PRIOR FILING DATE: 2001-03-09
PRIOR APPLICATION NUMBER: 60/280982
PRIOR FILING DATE: 2001-04-03
PRIOR APPLICATION NUMBER: 60/282129
PRIOR FILING DATE: 2001-04-04
PRIOR APPLICATION NUMBER: 60/282199
PRIOR FILING DATE: 2001-04-04
PRIOR APPLICATION NUMBER: 60/290589
PRIOR FILING DATE: 2001-05-09

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 18:19:54 ; Search time 8.3333 Seconds
(without alignments)
103.862 Million cell updates/sec

Title: US-10-014-658-8

Perfect score: 46

Sequence: 1 SEEGSGR 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 76:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39	84.8	402	2 T14438	phosphate/triose-phosphate translocator precursor, non-green plastid - wild cabbage
2	38	82.6	509	2 S18872	legumin-like storage protein - white spruce
3	37	80.4	1260	2 T26747	hypothetical prote
4	37	80.4	1276	2 C85899	protein Y39A1B.3 (
5	36	78.3	260	1 VCCVWV	coat protein - whe
6	36	78.3	260	2 S49386	capsid protein V2
7	36	78.3	447	2 A10701	probable amino aci
8	36	78.3	502	2 F31751	protein kinase cat
9	35	76.1	129	2 AC0782	probable DNA-bind
10	35	76.1	259	1 YXBYT	thymidylate syntha
11	35	76.1	287	2 B82239	heat shock protein
12	35	76.1	304	2 S44801	F10E9.2 protein.
13	35	76.1	319	2 H86218	protein F22QJ3.12
14	35	76.1	364	2 T00717	hypothetical prote
15	35	76.1	782	2 T32155	hypothetical prote
16	35	76.1	783	1 A38637	Ras interactor RIN
17	35	76.1	1089	2 T30843	serine-repeat anti
18	35	76.1	1232	2 B38432	ATP-dependent deox
19	35	76.1	147	2 T43502	hypothetical prote
20	34	73.9	246	2 B83418	DNA polymerase III
21	34	73.9	323	2 S78113	aldehyde reductase
22	34	73.9	327	2 T30072	hypothetical prote
23	34	73.9	359	2 T35179	vanillate O-demeth
24	34	73.9	389	2 T04816	myb-related protei
25	34	73.9	767	2 S55618	hypothetical prote
26	34	73.9	1827	2 JE0336	canalicul mulis
27	34	73.9	1898	1 A45973	crichonolal - num
28	33	71.7	154	2 A81184	hypothetical prote
29	33	71.7	154	2 F81917	hypothetical prote

ALIGNMENTS

RESULT 1

T14438

phosphate/triose-phosphate translocator precursor, non-green plastid - wild cabbage
C/Species: Brassica oleracea (wild cabbage)

C/Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 26-May-2000

C/Accession: T14438

R/Fischer, K./Weber, A.; Fluegge, U.I.

submitted to the EMBL Data Library, August 1994

A/Description: A triose-phosphate translocator from Cauliflower bud non-green plastids

A/Reference number: Z18089

A/Accession: T14438

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-402 <FIS>

A/Cross-references: EMBL:U13632; NID:g1143712; PID:g1143713

C/Genetics:

A/Gene: NCRPT

C/Superfamily: triose phosphate/3-phosphoglycerate/phosphate translocator

C/Keywords: plastid

F/1-82/Domain: transit peptide (non-green plastid) #status predicted <TNP>

F/83-402/Product: phosphate/phosphoenolpyruvate translocator #status predicted <MAT>

Query Match 84.8%; Score 39; DB 2; Length 402;

Best Local Similarity 87.5%; Pred. No. 17;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 EECEGSGR 9

|||||:

Db 88 EECEGSGK 95

RESULT 2

S18872

legumin-like storage protein - white spruce

C/Species: Picea glauca (white spruce)

C/Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 20-Aug-1999

C/Accession: S18872

R/Newton, C.H.

submitted to the EMBL Data Library, November 1991

A/Reference number: S18871

A/Accession: S18872

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-509 <NEW>

A/Cross-references: EMBL:X63192; NID:g20498; PIDN:CAA44874.1; PID:g20499

C/Superfamily: glycinin

Query Match 82.6%; Score 38; DB 2; Length 509;

Best Local Similarity 87.5%; Pred. No. 33;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 EECEGSGR 9

telokin - rabbit
squamosa promoter
squamosa promoter
hypothetical prote
hypothetical prote
hfas protein limpo
hypothetical prote
hypothetical prote
bexd protein - hae
ribosomal protein
ribosomal protein
ribosomal protein
ribosomal protein
3-oxoacyl-(acyl-ca
hypothetical prote
hypothetical prote
hypothetical prote

33 71.7 155 2 A41675
33 71.7 167 2 T52599
33 71.7 174 2 T52600
33 71.7 196 2 S08677
33 71.7 226 2 T08335
33 71.7 337 2 A87575
33 71.7 376 2 T48714
33 71.7 394 1 BMHIXD
33 71.7 396 1 RXLLIA
33 71.7 416 2 G86341
33 71.7 421 1 JC4277
33 71.7 426 1 T09551
33 71.7 426 2 D87457
33 71.7 432 2 T66639
33 71.7 461 2 H83882
33 71.7 492 2 A87471

```

RESULT 8
P31751
protein kinase catalytic chain homolog DC2 - fruit fly (Drosophila sp.)
C/Species: Drosophila sp.
C/Date: 21-May-1990 #sequence_revision 21-May-1990 #text_change 18-Jun-1999
C/Accession: F31751
R/Kalderon, D.; Rubin, G.M.
Genes Dev. 2, 1539-1556, 1988
A/Title: Isolation and characterization of Drosophila cAMP-dependent protein kinase gene
A/Reference number: A31751; MUID:89107990; PMID:3215511
A/Accession: F31751
A/Molecule type: mRNA
A/Residues: 1-502 <KAL>
A/Cross-references: EMBL:X16961; NID:G7815; PIDN:CAA34835.1; PID:G7816
C/Genetics:
A/Gene: FlyBase:PkA-C3
A/Cross-references: FlyBase:FBgn0000489
C/Superfamily: Kinase-related transforming protein; protein kinase homology
C/Keywords: ATP; phosphotransferase; serine/threonine-specific protein kinase
F/191-447/Domain: protein kinase homology <KIN>
F/199-207/Region: protein kinase ATP-binding motif

Query Match 78.3%; Score 36; DB 2; Length 502;
Best Local Similarity 75.0%; Pred. No. 75;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 BEGEGSGR 9
DB 136 DEEGNGR 143

RESULT 9
AC0782
probable DNA-binding protein STY2429 [imported] - Salmonella enterica subsp. enterica ser
C/Species: Salmonella enterica subsp. enterica serovar Typhi
A/Note: This species has also been called Salmonella typhi
C/Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C/Accession: AC0782
R/Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A/Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A/Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A/Reference number: AB0502; MUID:21534947; PMID:11677608
A/Accession: AC0782
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-129 <PAR>
A/Cross-references: GB:AL513382; PIDN:CAD02577.1; PID:G16503435; GSPDB:GN00176
C/Genetics:
A/Gene: STY2429

Query Match 76.1%; Score 35; DB 2; Length 129;
Best Local Similarity 87.5%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 BEGEGSGR 9
DB 88 EGEGSGR 95

RESULT 10
XYBYT
thymidylate synthase (EC 2.1.1.45) - yeast (Saccharomyces cerevisiae)
N/Alternate names: protein O2950; protein YOR074C
C/Species: Saccharomyces cerevisiae
C/Date: 31-Mar-1993 #sequence_revision 13-Mar-1997 #text_change 16-Jun-2000
C/Accession: S66957; A29546
R/Bohn, C.; Bolotin-Fukuhara, M.; Daignan-Fornier, B.; Dang, D.V.; Valens, M.
submitted to the Protein Sequence Database, July 1996
A/Reference number: S66929
A/Accession: S66957

```

```

A/Molecule type: DNA
A/Residues: 1-259 <BOH>
A/Cross-references: EMBL:Z74982; NID:G1420228; PIDN:CAA99267.1; PID:G1420229; GSPDB:GN0
A/Experimental source: strain S288C
R/Taylor, G.R.; Lajosky, P.A.; Storms, R.K.; Haynes, R.H.
J. Biol. Chem. 262, 5298-5307, 1987
A/Title: Molecular characterization of the cell cycle-regulated thymidylate synthase ge
A/Reference number: A29546; MUID:87165970; PMID:3031048
A/Accession: A29546
A/Molecule type: DNA
A/Residues: 1-33; 'GTLSLFAPPQRLRSLRDTFPLTTKKVFRGILLLWFLAGT', 34-259 <TAY>
A/Cross-references: GB:J02706; NID:G172989; PIDN:AAA60940.1; PID:G172990
C/Genetics:
A/Gene: SGD:CD21; TMD1; MIPS:YOR074C
A/Cross-references: SGD:S0005600; MIPS:YOR074C
A/Map position: 15R
A/Introns: 34/1
C/Superfamily: thymidylate synthase; thymidylate synthase homology
C/Keywords: deoxyribonucleotide biosynthesis; homodimer; methyltransferase; nucleus; py
F/10-259/Domain: thymidylate synthase homology <TDS>
F/132/Active site: Cys #status predicted

Query Match 76.1%; Score 35; DB 1; Length 259;
Best Local Similarity 75.0%; Pred. No. 59;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 BEGEGSGR 9
DB 144 KEEGSGK 151

RESULT 11
E82239
heat shock protein HtpX VC1117 [imported] - Vibrio cholerae (strain N16961 serogroup O1
C/Species: Vibrio cholerae
C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C/Accession: E82239
R/Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers,
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 405, 477-483, 2000
A/Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A/Reference number: A82035; MUID:20406833; PMID:10952301
A/Accession: E82239
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-287 <HEI>
A/Cross-references: GB:AE004192; GB:AE003852; NID:G9655581; PIDN:AAF94276.1; GSPDB:GN00
A/Experimental source: serogroup O1; strain N16961; biotype El Tor
C/Genetics:
A/Gene: VC1117
A/Map position: 1
C/Superfamily: heat-shock protein htpX

Query Match 76.1%; Score 35; DB 2; Length 287;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SEEGEGS 7
DB 183 SEEGEGS 189

RESULT 12
S44801
F10E9.2 protein - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 14-Sep-1994 #sequence_revision 12-May-1995 #text_change 31-Mar-2001
C/Accession: S44801
R/Anderson, K.
submitted to the EMBL Data Library, February 1993
A/Description: Sequence of the C. elegans cosmid F10E9.
A/Reference number: S44777

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 16:54:27 ; Search time 4.6667 Seconds
(without alignments)
90.694 Million cell updates/sec

Title: US-10-014-658-8
Perfect score: 46
Sequence: 1 SEEGEGSGR 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39	84.8	402	1 CPT2 BRAOL	P52178 brassica ol
2	36	78.3	260	1 COAT WDV	P06946 wheat dwarf
3	36	78.3	502	1 KDC2 DRONE	P16912 drosophila
4	35	76.1	286	1 YLU2 CAHEL	P34396 caenorhabdi
5	35	76.1	287	1 HTPX VIBCH	Q9KSY9 vibrio chol
6	35	76.1	304	1 TFSY YEAST	P06785 saccharomyc
7	35	76.1	655	1 NOG1 TRYEB	Q9U6A9 trypanosoma
8	35	76.1	783	1 RIN1 HUMAN	Q13671 homo sapien
9	35	76.1	889	1 ENV STVSP	P19503 simian immu
10	35	76.1	1189	1 HAIR HUMAN	O43593 homo sapien
11	35	76.1	1232	1 ADDA BACSU	P23478 bacillus su
12	34	73.9	266	1 HIC2 MOUSE	Q9J126 mus musculus
13	34	73.9	322	1 ALDX SPOSA	P27800 sporobolomy
14	34	73.9	1048	1 AGO1 ARATH	O04379 arabidopsis
15	34	73.9	1191	1 SMC3 RAT	P97690 rattus norv
16	34	73.9	1217	1 SMC3 HUMAN	Q9UQE7 homo sapien
17	34	73.9	1217	1 SMC3 MOUSE	Q9CW03 mus musculus
18	34	73.9	1527	1 MRP3 HUMAN	O15438 homo sapien
19	34	73.9	1898	1 TRHY HUMAN	Q07283 homo sapien
20	34	73.9	8797	1 SNE1 HUMAN	Q8N191 homo sapien
21	33	71.7	329	1 AXIC ARATH	O22048 arabidopsis
22	33	71.7	337	1 HPAB CAUCR	P27343 caulobacter
23	33	71.7	394	1 BEND HAEN	P22336 haemophilus
24	33	71.7	396	1 RL4 XENLA	P08429 xenopus lae
25	33	71.7	419	1 RL4 MOUSE	Q9D8E6 mus musculus
26	33	71.7	420	1 RL4 CANFA	Q28346 canis fami
27	33	71.7	421	1 RL4 RAT	P50878 rattus norv
28	33	71.7	427	1 RL4 HUMAN	P36578 homo sapien
29	33	71.7	438	1 KMLS SHEEP	O02827 ovis aries
30	33	71.7	629	1 DXS DEIRA	Q9RUB5 deinococcus
31	33	71.7	796	1 YPC3 YEAST	P43573 saccharomyc
32	33	71.7	940	1 LON2 ARATH	P36555 arabidopsis
33	33	71.7	1147	1 KMLS_RABIT	P29294 coryctolagus

34	33	71.7	1176	1 KMLS BOVIN	Q28824 bos taurus
35	33	71.7	1914	1 KMLS_HUMAN	Q15746 homo sapien
36	33	71.7	2280	1 YCF2_OENHO	Q9NEF2 oenothera h
37	33	71.7	2280	1 YCF2_TOBAC	P09766 nicotiana t
38	33	71.7	2294	1 YCF2_ARATH	P56786 arabidopsis
39	32	69.6	174	1 NU6M_GORGO	Q34573 gorilla gor
40	32	69.6	174	1 NU6M_HUMAN	P03923 homo sapien
41	32	69.6	174	1 NU6M_HYLLA	Q95710 hylobates l
42	32	69.6	174	1 NU6M_PAPHA	Q9ZXX5 papio hamad
43	32	69.6	174	1 NU6M_PONPA	P32700 pongo pygma
44	32	69.6	205	1 DNEI_ADELI2	P36704 human adeno
45	32	69.6	206	1 KTHY_CHLPN	Q9Z8R5 chlamydia p

ALIGNMENTS

RESULT 1					
CPT2 BRAOL					
ID	CPT2 BRAOL	STANDARD;	PRT;	402 AA.	
AC	P52178;				
DT	01-OCT-1996 (Rel. 34, Created)				
DT	01-OCT-1996 (Rel. 34, Last sequence update)				
DT	28-FEB-2003 (Rel. 41, Last annotation update)				
DE	Triose phosphate/phosphate translocator, non-green plastid,				
DE	Chloroplast precursor (CTPT).				
GN	NGTPT.				
OS	Brassica oleracea (Cauliflower).				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;				
OC	euroside II; Brassicales; Brassicaceae; Brassica.				
OX	NCBI_TaxID=3712;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=cv. Botrytis; TISSUE=Flower;				
RX	MEDLINE=97246325; PubMed=9090886;				
RA	Fischer K., Kammerer B., Gutenschn M., Arbinger B., Weber A.,				
RA	Kausler R.E., Plugge U.-I.;				
RT	"A new class of plastidic phosphate translocators: a putative link				
RT	between primary and secondary metabolism by the				
RT	phosphoenolpyruvate/phosphate antiporter.";				
RL	Plant Cell 9:453-462(1997).				
CC	-!- FUNCTION: MEDIATES THE EXPORT OF FIXED CARBONS FROM THE				
CC	CHLOROPLASTS INTO THE CYTOSOL IN THE FORM OF TRIOSE PHOSPHATES.				
CC	-!- SUBUNIT: Homodimer (By similarity).				
CC	-!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. LOCATED IN				
CC	ZONES OF CONTACT BETWEEN THE INNER AND OUTER MEMBRANE OF THE				
CC	CHLOROPLAST (BY SIMILARITY).				
CC	-!- SIMILARITY: STRONG, TO OTHER PLANTS CPTP. ALSO SIMILAR TO				
CC	YEAST PROTEIN SLX41.				

This SWISS-PROT entry is copyright. It is produced through a collaboration					
between the Swiss Institute of Bioinformatics and the EMBL outstation -					
the European Bioinformatics Institute. There are no restrictions on its					
use by non-profit institutions as long as its content is in no way					
modified and this statement is not removed. Usage by and for commercial					
entities requires a license agreement (See http://www.isb-sib.ch/announce/					
or send an email to license@isb-sib.ch).					

EMBL; U13632; AAA84892.1; --					
PIR; T14438; T14438.					
InterPro; IPR000620; DUF6.					
InterPro; IPR004696; Tpt_PEP_transl.					
Pfam; PF00892; DUF6; 1.					
TIGRfams; TIGR00817; tpt; 1.					
Transmembrane; Chloroplast;					
TRANSIT	1	82			
CHAIN	83	402			
DOMAIN	83	99			
TRANSMEM	99	119			
DOMAIN	120	131			
TRANSMEM	132	152			

Fri Dec 13:21:54 2003

us-10-01 18-1.rpr

Page 5

Search completed: December 11, 2003, 18:29:12
Job time : 14.3333 secs

probable kinase [imported] - Mycobacterium leprae
 C:Species: Mycobacterium leprae
 C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
 C:Accession: G97133
 R:Coate, S.T.; Eigmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Hc
 R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
 eam, M.A.; Rutherford, K.M.
 Nature 409, 1007-1011, 2001
 A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq
 A:Title: Massive gene decay in the leprosy bacillus.
 A:Reference number: A96909; MUID:21128732; PMID:11234002
 A:Accession: G97133
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-327 <STO>
 A:Cross-references: GB:AL450380; NID:G1093510; PIDN:CAC30751.1; GSPDB:GN00147
 C:Genetics:
 A:Gene: ML1798

Query Match 78.0%; Score 32; DB 2; Length 327;
 Best Local Similarity 75.0%; Pred. No. 28;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 STALEAIG 8
 |||:|:|:
 Db 68 STALEALG 75

RESULT 8
 C70712
 Hypothetical protein Rv1496 - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
 C:Accession: C70712
 R:Coate, S.T.; Brosch, R.; Parkhill, J.; Garnier, C.; Churcher, C.; Harris, D.; Gordon, S
 ; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S
 ; Rajandream, M.A.; Rogers, J.; Rutter, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A:Reference number: A70500; MUID:98295987; PMID:9634230
 A:Accession: C70712
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-334 <COL>
 A:Cross-references: GB:T9701; GB:AL123456; NID:G3261635; PIDN:CAB02046.1; PID:ex64125;
 A:Experimental source: strain H37RV
 C:Genetics:
 A:Gene: Rv1496

Query Match 78.0%; Score 32; DB 2; Length 334;
 Best Local Similarity 75.0%; Pred. No. 29;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 STALEAIG 8
 |||:|:|:
 Db 72 STALEALG 79

RESULT 9
 G96027
 Probable 3-carboxy-cis-cis-mucnate cycloisomerase (EC 5.5.1.2) [imported] - Sinorhizobi
 C:Species: Sinorhizobium meliloti
 C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 14-Sep-2001
 C:Accession: G96027
 R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan
 Proc. Natl. Acad. Sci. U.S.A. 98, 9899-9894, 2001
 A:Title: The complete sequence of the 1.683-kb pSymB megaplasmid from the N2-fixing endo
 A:Reference number: A95842; MUID:21396508; PMID:11481431
 A:Accession: G96027
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-351 <KUR>

A:Cross-references: GB:AL591985; PIDN:CAC49887.1; PID:gl5141375; GSPDB:GN00167
 A:Experimental source: strain 1021, megaplasmid pSymb
 R:Galibert, F.; Finan, T.M.; Long, S.R.; Fuhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler
 pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Pederspiel, N.A.; Fisher, R.P.
 L.; Hyman, R.W.; Jones, T.
 Science 293, 668-672, 2001
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure
 heault, P.; Vandenberg, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K
 A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
 A:Reference number: A96039; MUID:21368234; PMID:11474104
 A:Contents: annotation
 C:Genetics:
 A:Gene: pcaB; SMb20575
 A:Genome: Plasmid
 C:Keywords: Intramolecular lyase; isomerase

Query Match 78.0%; Score 32; DB 2; Length 351;
 Best Local Similarity 87.5%; Pred. No. 30;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TALEAIGR 9
 |||:|:|:
 Db 136 TALEACGR 143

RESULT 10
 AE1397
 conserved hypothetical protein lmo2581 [imported] - Listeria monocytogenes (strain EGD-1
 C:Species: Listeria monocytogenes
 C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
 C:Accession: AE1397
 R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecke
 ; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H
 D.; Jones, L.M.; Karst, U.
 Science 294, 849-852, 2001
 A:Authors: Krefst, J.; Kahn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; M
 ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland
 A:Title: Comparative genomics of Listeria species.
 A:Reference number: AB1077; MUID:21537279; PMID:11679669
 A:Accession: AE1397
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-362 <GLA>
 A:Cross-references: GB:NC_003210; PIDN:CAD00659.1; PID:gl6412069; GSPDB:GN00177
 A:Experimental source: strain EGD-e
 C:Genetics:
 A:Gene: lmo2581

Query Match 78.0%; Score 32; DB 2; Length 362;
 Best Local Similarity 100.0%; Pred. No. 31;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 ALBAIGR 9
 |||:|:|:
 Db 354 ALBAIGR 360

RESULT 11
 AH1772
 conserved hypothetical protein lin2726 [imported] - Listeria innocua (strain Clip11262)
 C:Species: Listeria innocua
 C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
 C:Accession: AH1772
 R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecke
 ; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H
 D.; Jones, L.M.; Karst, U.
 Science 294, 849-852, 2001
 A:Authors: Krefst, J.; Kahn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; M
 ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland
 A:Title: Comparative genomics of Listeria species.
 A:Reference number: AB1077; MUID:21537279; PMID:11679669
 A:Accession: AH1772
 A:Status: preliminary

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 18:19:54 ; Search time 8.3333 Seconds
(without alignments)
103.862 Million cell updates/sec

Title: US-10-014-658-1
Perfect score: 41
Sequence: 1 STALEAIGR 9

Scoring table: BLOSUMP62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 76: *
1: PIR1: *
2: PIR2: *
3: PIR3: *
4: PIR4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	37	90.2	535	2 S07521	head-to-tail joini
2	37	90.2	536	1 JOBPT7	head-to-tail joini
3	33	80.5	621	2 T18737	hypothetical prote
4	33	80.5	702	2 A69140	ATP-dependent heli
5	32	78.0	261	2 E95563	probable transcrip
6	32	78.0	293	2 AB1878	cyanoophycinase [i
7	32	78.0	327	2 G87133	probable kinase [i
8	32	78.0	334	2 C70712	hypothetical prote
9	32	78.0	351	2 G96037	probable 3-carboxy
10	32	78.0	362	2 AB1397	conserved hypothet
11	32	78.0	362	2 AH1772	conserved hypothet
12	32	78.0	393	2 F84068	aspartate aminotra
13	32	78.0	480	2 T06048	xanthine/uracil pe
14	32	78.0	744	2 T06048	kinesin-related pr
15	32	78.0	754	2 S48020	hypothetical prote
16	32	78.0	896	2 T24169	hypothetical prote
17	32	78.0	906	2 T24166	probable reverse t
18	32	78.0	1206	2 S72620	probable guanine n
19	32	78.0	1313	2 T38943	resistance protein
20	32	78.0	1804	2 T30553	resistance protein
21	32	78.0	1810	2 T30562	resistance protein
22	32	78.0	1813	2 T30564	resistance protein
23	32	78.0	1847	2 T30558	resistance transcrip
24	31	75.6	300	2 D83502	probable biosynthes
25	31	75.6	311	2 D32786	thiamin biosynthes
26	31	75.6	318	2 T36282	probable dehydroge
27	31	75.6	322	1 JQ1028	6-phosphofructokin
28	31	75.6	367	2 G95948	probable glycosylc
29	31	75.6	426	2 AE2635	polysaccharide bio

30	31	75.6	430	2 D97417	hypothetical prote
31	31	75.6	602	2 T45278	oligopeptide ABC t
32	31	75.6	1107	2 T25450	hypothetical prote
33	31	75.6	2335	2 T40186	probable phosphati
34	31	75.6	3085	2 T00327	polyprotein - infe
35	30	73.2	95	1 WTRBM1	macrophage antibio
36	30	73.2	95	1 WTRBM2	defensin CS-4 prec
37	30	73.2	150	2 C72697	hypothetical prote
38	30	73.2	177	2 T45393	hypothetical prote
39	30	73.2	233	2 D84239	probable triosepho
40	30	73.2	249	2 F71311	hypothetical prote
41	30	73.2	263	2 C84397	conserved hypothet
42	30	73.2	266	2 E83429	acetyltransferase,
43	30	73.2	312	2 E87350	DNA repair protein
44	30	73.2	327	2 F84290	hypothetical prote
45	30	73.2	337	2 E95913	

ALIGNMENTS

RESULT 1

S07521
head-to-tail joining protein - phage T3
N/Alternate names: Head-tail connector protein
C/Species: phage T3
C/Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 20-Sep-1999
C/Accession: S07521
R/Back, P.J.; Gonzalez, S.; Ward, C.L.; Molineux, I.J.
J. Mol. Biol. 210, 687-701, 1989
A/Title: Sequence of bacteriophage T3 DNA from gene 2.5 through gene 9.
A/Reference number: S07500; MUID:90133923; PMID:2614843
A/Accession: S07521
A/Molecule type: DNA
A/Residues: 1-535 <BEC>
A/Cross-references: EMBL:X17255; NID:g15682; PIDN:CAA35152.1; PID:g15714
C/Genetics:
A/Gene: 8
C/Superfamily: phage T7 head-to-tail joining protein

Query Match 90.2%; Score 37; DB 2; Length 535;
Best Local Similarity 88.9%; Pred. No. 3.5;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 STALEAIGR 9
DB 427 STGLEAIGR 435

RESULT 2

JOBPT7
head-to-tail joining protein - phage T7
C/Species: phage T7
C/Date: 13-Jun-1983 #sequence_revision 13-Jun-1983 #text_change 23-Jul-1999
C/Accession: A04354; S42323
R/Dunn, J.J.; Thompson, K.
submitted to the Nucleic Acid Sequence Database, September 1982
A/Reference number: A94615
A/Accession: A04354
A/Molecule type: DNA
A/Residues: 1-536 <DUN>
R/Dunn, J.J.; Studier, F.W.
J. Mol. Biol. 166, 477-535, 1983
A/Title: Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 91
A/Reference number: S42283; MUID:83241725; PMID:6864790
A/Accession: S42323
A/Molecule type: DNA
A/Residues: 1-536 <DUN>
A/Cross-references: EMBL:V01146; NID:g431187; PIDN:CAA44425.1; PID:g15602
A/Note: the authors did not translate the codon for residue 1
C/Genetics:
A/Gene: 8
A/Map position: 50.68-54.70

DR SMART; SMO0485; XPGN; 1.
 DR PROSITE; PS00841; XPG 1; FALSE NEG.
 KW Hydrolyase; Nuclease; Endonuclease; Magnesium; Metal-binding;
 KW Complete proteome. 156 MAGNESIUM 1 (BY SIMILARITY).
 FT METAL 156
 SQ SEQUENCE 327 AA; 35814 MW; 7D597B212864D549 CRC64;

Query Match 73.2%; Score 30; DB 1; Length 327;
 Best Local Similarity 75.0%; Pred. No. 50;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 STALEAIG 8
 DB 237 TTDALDAIG 244

RESULT 15
 ID GBBI DROME STANDARD; PRT; 340 AA.
 AC P26308; Q9VXV8;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Guanine nucleotide-binding protein beta subunit 1.
 GN G-BETA-13F OR GB13F OR CG10545.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89017152; PubMed=3140235;
 RA Yarifitz S., Provost N.M., Hurley J.B.;
 RT "Cloning of a Drosophila melanogaster guanine nucleotide regulatory
 RT protein beta-subunit gene and characterization of its expression
 RT during development.";
 RL Proc. Natl. Acad. Sci. U.S.A. 85:7134-7138(1988).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
 RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,
 RA Lasko P., Lei Y., Levitsky A.A., Li Z., Li Z., Liang X., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J., Yao Q.A.,
 Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 RN [3]
 RN SEQUENCE FROM N.A.
 RP STRAIN=Berkeley; TISSUE=Embryo;
 RX MEDLINE=22426066; PubMed=12537569;
 RA Stapleton M., Carlson J.W., Brockstein P., Yu C., Champe M.,
 RA George R.A., Guarin H., Krommiller B., Pacle J.M., Park S., Wan K.H.,
 RA Rubin G.M., Ceiniker S.E., CDNA resource.";
 RL "A Drosophila full-length cDNA resource.";
 CC Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002)
 CC -!- FUNCTION: GUANINE NUCLEOTIDE-BINDING PROTEINS (G PROTEINS) ARE
 CC INVOLVED AS A MODULATOR OR TRANSDUCER IN VARIOUS TRANSMEMBRANE
 CC SIGNALING SYSTEMS. THE BETA AND GAMMA CHAINS ARE REQUIRED FOR THE
 CC GTPASE ACTIVITY, FOR REPLACEMENT OF GDP BY GTP, AND FOR G PROTEIN-
 CC EFFECTOR INTERACTION.
 CC -!- SUBUNIT: G proteins are composed of 3 units (alpha, beta and
 CC gamma).
 CC -!- SIMILARITY: Contains 7 WD repeats.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M22567; AAB59247.1; -;
 DR EMBL; AB003500; AAF48530.1; -;
 DR EMBL; AV058566; AAL13795.1; -;
 DR PIR; A40489; RGFBH.
 DR HSP; P04901; ITRG.
 DR FlyBase; FBgn0011105; G-beta-13F.
 DR InterPro; IPR001632; Gprotein_B.
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF00400; WD40; 7.
 DR PRINTS; PR00319; GPROTEINB.
 DR PRINTS; PR00320; GPROTEINRPT.
 DR ProDom; PD000018; WD40; 4.
 DR SMART; SM00320; WD40; 7.
 DR PROSITE; PS00678; WD_REPEATS_1; 3.
 DR PROSITE; PS00082; WD_REPEATS_2; 6.
 DR PROSITE; PS00294; WD_REPEATS_REGION; 1.
 KW Transducer; Repeat; WD repeat; Multigene family.
 FT REPEAT 53 83
 FT REPEAT 95 125 WD 2.
 FT REPEAT 141 170 WD 3.
 FT REPEAT 182 212 WD 4.
 FT REPEAT 224 254 WD 5.
 FT REPEAT 268 298 WD 6.
 FT REPEAT 310 340 WD 7.
 SQ SEQUENCE 340 AA; 37133 MW; 5A4B70DCB0A1C32C CRC64;

Query Match 73.2%; Score 30; DB 1; Length 340;
 Best Local Similarity 56.7%; Pred. No. 52;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 STALEAIG 9
 DB 34 ATSLPIGR 42

Search completed: December 11, 2003, 18:20:50
 Job time : 16.6667 secs

```

RX TISSUE=Peritoneal neutrophils;
RA MEDLINE=85182561; PubMed=3988726;
RA Selsted M.E., Brown D.M., Delange R.J., Harwig S.S.L., Lehrer R.I.;
RT "Primary structures of six antimicrobial peptides of rabbit
RT peritoneal neutrophils.";
RL J. Biol. Chem. 260:4579-4584(1985).
RN [4]
RP SEQUENCE OF 63-95.
RC TISSUE=Lung;
RX MEDLINE=92164536; PubMed=1311240;
RA Zhu Q., Solomon S.;
RT "Isolation and mode of action of rabbit corticostatic
RT (antidrenocorticotropin) peptides.";
RL Endocrinology 130:1413-1423(1992).
CC -!- FUNCTION: THIS PEPTIDE HAS ANTI-BIOTIC, ANTI-FUNGI AND ANTIVIRAL
CC ACTIVITY. IT ALSO INHIBITS CORTICOTROPIN (ACTH) STIMULATED
CC CORTICOSTERONE PRODUCTION.
CC -!- SIMILARITY: BELONGS TO THE CORTICOSTATIN/DEFENSIN FAMILY.
CC -!- SIMILARITY: THERE ARE TWO DIFFERENCES BETWEEN THE PRECURSORS OF
CC MCP-1 AND OF MCP-2: ONE IN THE PROPEPTIDE, AND ONE IN THE MATURE
CC SEQUENCE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M28883; AAA31387.1; -
DR EMBL; M28072; AAA31388.1; -
DR PIR; A45811; WTRM1.
DR InterPro; IPR006081; Defensin_alpha.
DR InterPro; IPR006080; Defensin_mammal.
DR InterPro; IPR002366; Defensin_propep.
DR Pfam; PF00879; Defensin_propep; 1.
DR Pfam; PF00323; defensins; 1.
DR SMART; SM00048; DFESN; 1.
DR PROSITE; PS00249; DEFENSIN; 1.
DR Defensin; Antibioc; Antiviral; Fungicide; Signal.
KW SIGNAL
FT SIGNAL 1 19 POTENTIAL.
FT PROPEP 20 62
FT CHAIN 63 95 CORTICOSTATIN III.
FT DISULPID 65 93 BY SIMILARITY.
FT DISULPID 67 82 BY SIMILARITY.
FT DISULPID 72 92 BY SIMILARITY.
FT DISULPID 72 92 BY SIMILARITY.
SQ SEQUENCE 95 AA; 10460 MW; 56593721BB6EF032 CRC64;

Query Match 73.2%; Score 30; DB 1; Length 95;
Best Local Similarity 75.0%; Pred. No. 15;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 STALEAIG 8
Db |:||||:|
51 SSALEALG 58

RESULT 12
DEF4 RABIT STANDARD; PRT; 95 AA.
AC P01377;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Corticostatin IV precursor (CS-IV) (Macrophage antibiotic peptide MCP-
DE 2) (NP-2) (Antidrenocorticotropin peptide IV).
OS Oryctolagus cuniculus (Rabbit)
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
CX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=89309825; PubMed=2745983;
RA Ganz T., Rayner J.R., Valore E.V., Tumolo A., Talmadge K., Fuller F.;
RT "The structure of the rabbit macrophage defensin genes and their
RT organ-specific expression.";
RL J. Immunol. 143:1358-1365(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93176141; PubMed=8439302;
RA Sadro L.C., Tremblay A., Solomon S., Palfree R.G.E.;
RT "Differential expression of corticostatin/defensins: higher levels
RT of CS-4 (NP-2) transcripts compared with CS-6 (NP-5) in rabbit
RT lung.";
RL Biochem. Biophys. Res. Commun. 190:1009-1016(1993).
RN [3]
RP SEQUENCE OF 63-95.
RC TISSUE=Lung macrophage;
RX MEDLINE=84061901; PubMed=6643497;
RA Selsted M.E., Brown D.M., Delange R.J., Lehrer R.I.;
RT "Primary structures of MCP-1 and MCP-2, natural peptide antibiotics
RT of rabbit lung macrophages.";
RL J. Biol. Chem. 268:14485-14489(1993).
RN [4]
RP SEQUENCE OF 63-95.
RC TISSUE=Peritoneal neutrophils;
RX MEDLINE=85182561; PubMed=3988726;
RA Selsted M.E., Brown D.M., Delange R.J., Harwig S.S.L., Lehrer R.I.;
RT "Primary structures of six antimicrobial peptides of rabbit
RT peritoneal neutrophils.";
RL J. Biol. Chem. 260:4579-4584(1985).
RN [5]
RP SEQUENCE OF 63-95.
RC TISSUE=Lung;
RX MEDLINE=92164536; PubMed=1311240;
RA Zhu Q., Solomon S.;
RT "Isolation and mode of action of rabbit corticostatic
RT (antidrenocorticotropin) peptides.";
RL Endocrinology 130:1413-1423(1992).
RN [6]
RP STRUCTURE BY NMR OF 63-95.
RX MEDLINE=93075733; PubMed=1445872;
RA Zhang X.-L., Selsted M.E., Pardi A.;
RT "NMR studies of defensin antimicrobial peptides. 1. Resonance
RT assignment and secondary structure determination of rabbit NP-2 and
RL human HNP-1.";
RN Biochemistry 31:11348-11356(1992).
RN [7]
RP STRUCTURE BY NMR OF 63-95.
RX MEDLINE=93075734; PubMed=1445873;
RA Pardi A., Zhang X.-L., Selsted M.E., Skalicky J.J., Yip P.F.;
RT "NMR studies of defensin antimicrobial peptides. 2. Three-dimensional
RL structures of rabbit NP-2 and human HNP-1.";
RN Biochemistry 31:11357-11364(1992).
CC -!- FUNCTION: THIS PEPTIDE HAS ANTI-BIOTIC, ANTI-FUNGI AND ANTIVIRAL
CC ACTIVITY. IT ALSO INHIBITS CORTICOTROPIN (ACTH) STIMULATED
CC CORTICOSTERONE PRODUCTION.
CC -!- SIMILARITY: BELONGS TO THE CORTICOSTATIN/DEFENSIN FAMILY.
CC -!- SIMILARITY: THERE ARE TWO DIFFERENCES BETWEEN THE PRECURSORS OF
CC MCP-1 AND OF MCP-2: ONE IN THE PROPEPTIDE, AND ONE IN THE MATURE
CC SEQUENCE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M28884; AAA31390.1; -
DR EMBL; S55582; AAB25449.2; -
DR EMBL; M28073; AAA31389.1; -
DR EMBL; L10841; AAA31424.1; -
DR PIR; B45811; WTRM2.

```

```
DB 155 SKALEAICK 163
|||||:
RESULT 8
K6P1_TETH
ID K6P1_TETH STANDARD; PRT; 322 AA.
AC P21777
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-WAR-1992 (Rel. 21, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE 6-phosphofructokinase isozyme 1 (EC 2.7.1.11) (Phosphofructokinase-1)
DE (PFK1).
DE PFK1.
GN PFK1.
OS Thermus thermophilus.
OC Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
OC Thermus.
OX NCBI_TaxID=274;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HB8 / ATCC 27634;
RX MEDLINE=91246220; PubMed=1828151;
RA Xu J., Seki M., Denda K., Yoshida M.;
RT "Molecular cloning of phosphofructokinase 1 gene from a thermophilic
bacterium, Thermus thermophilus.";
RL Biochem. Biophys. Res. Commun. 176:1313-1318(1991).
RN [2]
RP SEQUENCE OF 1-25.
RC STRAIN=HB8 / ATCC 27634;
RX MEDLINE=91324297; PubMed=1830879;
RA Xu J., Oshima T., Yoshida M.;
RT "Phosphoenolpyruvate-insensitive phosphofructokinase isozyme from
Thermus thermophilus HB8.";
RL J. Biochem. 109:199-203(1991).
CC -1- CATALYTIC ACTIVITY: ATP + D-fructose 6-phosphate = ADP + D-
fructose 1,6-bisphosphate.
CC -1- ENZYME REGULATION: ALLOSTERICALLY INHIBITED BY PHOSPHOENOLPYRUVATE
WHICH INDUCES THE DISSOCIATION OF THE ACTIVE TETRAMER INTO AN
INACTIVE TWO-SUBUNIT FORMS.
CC -1- PATHWAY: Key control step of glycolysis.
CC -1- SUBUNIT: Homotetramer.
CC -1- SIMILARITY: Belongs to the phosphofructokinase family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; M71213; AAA27501.1; -
DR HAMAP; MF 00339; -; 1.
DR InterPro; IPR000023; Pfruckinase.
DR Pfam; PF00365; PFK; 1.
DR PRINTS; PR00476; PFRCTKINASE.
DR ProDom; PD000707; Pfruckinase; 1.
DR PROSITE; PS00433; PHOSPHOFRUCTOKINASE; 1.
KW Kinase; Transferase; Glycolysis; Allosteric enzyme; Multigene family.
SQ SEQUENCE 322 AA; 33606 MW; 92945F3B204A9D15 CRC64;
-----
Query Match 75.6%; Score 31; DB 1; Length 322;
Best Local Similarity 77.8%; Pred. No. 29;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 STALEAIGR 9
|:|:|:|:|
DB 145 NTALEAIDR 153
|:|:|:|:|
RESULT 9
COCH_CHICK
ID COCH_CHICK STANDARD; PRT; 547 AA.
AC 042163;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cochlin precursor (COCH-5B2).
GN COCH OR COCH5B2.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A. TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.
RC STRAIN=White leghorn; TISSUE=Basilar papilla;
RX MEDLINE=9849665; PubMed=9736749;
RA Heller S., Sheane C.A., Javed Z., Hudspeth A.J.;
RT "Molecular markers for cell types of the inner ear and candidate genes
for hearing disorders.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:11400-11405(1998).
CC -1- SUBCELLULAR LOCATION: Secreted (Potential).
CC -1- TISSUE SPECIFICITY: Expressed in inner ear structures; the
spindle-shaped cells of the basilar papilla. Weaker expression
found in the inferior and superior fibrocartilaginous plates and
skeletal muscle.
CC -1- DEVELOPMENTAL STAGE: Specifically expressed at the late
developmental stages in the cochlea.
CC -1- SIMILARITY: Contains 1 LCCL domain.
CC -1- SIMILARITY: Contains 2 VWFA domains.
CC -1- DATABASE: NAME=protein Spotlight;
NOTE=Issue 4 of November 2000;
WWW=http://www.expasy.org/spotlight/articles/spot004.html.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AF012252; AAC62253.1; -
DR HSSP; F11215; IJLM.
DR InterPro; IPR004043; LCCL_dom.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF03815; LCCL; 1.
DR Pfam; PF00092; vwa; 2.
DR PRINTS; PR00453; VWFADOMAIN.
DR SMART; SM00603; LCCL; 1.
DR SMART; SM00327; VWA; 2.
DR PROSITE; PS50820; LCCL; 1.
DR PROSITE; PS50234; VWFA; 2.
KW Glycoprotein; Repeat; Signal.
FT SIGNAL 1 19
FT CHAIN 20 547
FT DOMAIN 24 117
FT DOMAIN 162 347
FT DOMAIN 364 534
FT DOMAIN 50 46
FT DISULFID 30 46
FT DISULFID 50 70
FT CARBOHYD 218 218
SQ SEQUENCE 547 AA; 59426 MW; 172724242641DF88 CRC64;
-----
Query Match 75.6%; Score 31; DB 1; Length 547;
Best Local Similarity 85.7%; Pred. No. 50;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 3 ALEAIGR 9
|:|:|:|:|
DB 127 ALEAVGR 133
|:|:|:|:|
```

DB 354 TALEAIG 360

RESULT 5

ATK2 ARATH STANDARD; PRT; 745 AA.

AC P46864; Q9T047;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Kinesin 2 (Kinesin-like protein B).

GN ATK2 OR KATB OR Ar4G27180 OR 124A18.130.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.

OC NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;

RX MEDLINE=94355659; PubMed=8075402;

RA Mitsui H., Nakatani K., Yamaguchi-Shinozaki K., Shinozaki K.,

RA Nishikawa K., Takahashi H.,

RT "Sequencing and characterization of the kinesin-related genes katB

RT and katC of Arabidopsis thaliana."

RL Plant Mol. Biol. 25:865-876(1994).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;

RX MEDLINE=20083488; PubMed=10617198;

RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,

RA Pohl T., Duesterhoeft A., Striekema W., Entian K.-D., Terry N.,

RA Harris B., Anseuge W., Brandt P., Grivell L.A., Rieger M., Mueller M.,

RA Weichselgartner M., de Simone V., Obermaier B., Maché R., Schmidheini T.,

RA Kreis M., Delsen M., Puigdomenech P., Watson M., Schmidheini T.,

RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,

RA Vos P., Hoheisel J., Zimmermann W., Wedler H., Ridley P.,

RA Langham S.-A., McCullagh B., Bilham L., Robben J.,

RA Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,

RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,

RA Weitzenecker T., Bothe G., Rumpelberger U., Hilbert H., Braun M.,

RA Holzer E., Brandt P., Peters S., van Staveren M., Dircke W.,

RA Moolman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,

RA Berner S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,

RA De Keyser A., Buysschaert C., Gielen J., Villarroel R., De Clercq R.,

RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,

RA Clark L., Doggett J., Hall S., Kay M., Lennard N., Mclay K., Mayes R.,

RA Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,

RA Borkova D., Bloeker H., Schaefer M., Grimm M., Loenert T.-H.,

RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,

RA Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,

RA Neumann S., Argirion A., Vitale D., Liguori R., Piravandi E.,

RA Massenot O., Quigley F., Clabaud G., Muendlein A., Felber R.,

RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,

RA Chedof F., Cooke R., Berger C., Monfort A., Casacuberta E.,

RA Gibbons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A.,

RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,

RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielke C.,

RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,

RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,

RA Parnell M., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,

RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threiden U.,

RA Stonking T., Kallicki J., Graves T., Harmon G., Edwards J.,

RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,

RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,

RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,

RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman C.,

RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Jochu C.,

RA Antonciu B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,

RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,

RA Swaby I.K., O'Shaughnessy A., Rodriguez N., Hoffman J., Till S.,

RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,

RA Chen E., Marra M., Martienssen R., McCombie W.R.;

"Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.";

RT Nature 402:769-777(1999).

RL

CC -|- FUNCTION: POSSIBLE ROLE IN MITOSIS.

CC -|- DOMAIN: COMPOSED OF THREE STRUCTURAL DOMAINS: A SMALL GLOBULAR

CC N-TERMINAL, A CENTRAL ALPHA-HELICAL COILED COIL AND A LARGE

CC GLOBULAR C-TERMINAL WHICH IS RESPONSIBLE FOR THE MOTOR ACTIVITY

CC (IT HYDROLYZES ATP AND BINDS MICROTUBULES).

CC -|- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. NCD

CC SUBFAMILY.

CC -|- CAUTION: Ref.2 sequences differ from that shown due to erroneous

CC gene model prediction.

CC

CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

CC -----

CC EMBL; D211137; BAA04673.1; -

CC EMBL; AL035680; CAB38848.1; ALT SEQ.

CC EMBL; AL161566; CAB79573.1; ALT SEQ.

CC PIR; T06048; T06048.

CC HSP; P17119; 3KAR.

CC InterPro; IPR001752; kinesin_motor.

CC Pfam; PF00225; kinesin; 1.

CC PRINTS; PRO0380; KINESINHEAVY.

CC SMART; SMC0129; KISC; 1.

CC PROSITE; PS00411; KINESIN MOTOR DOMAIN1; 1.

CC PROSITE; PS00057; KINESIN MOTOR DOMAIN2; 1.

CC Motor protein; Microtubules; ATP-binding; Coiled coil;

CC Multigene family.

CC DOMAIN 1 35 GLOBULAR.

CC DOMAIN 36 384 COILED COIL.

CC DOMAIN 385 718 KINESIN-MOTOR (BY SIMILARITY).

CC FT NP BIND 472 479 ATP (POTENTIAL).

CC FT NP BIND 745 AA; 84359 MW; 6DB352FDF4FD7BC1 CRC64;

CC SQ SEQUENCE 745 AA; 84359 MW; 6DB352FDF4FD7BC1 CRC64;

Query Match 78.0%; Score 32; DB 1; Length 745;

Best Local Similarity 75.0%; Pred. No. 41;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QV 2 TALEAIGR 9

Db 414 TSLEALGR 421

RESULT 6

ATK3 ARATH STANDARD; PRT; 754 AA.

ID ATK3 ARATH STANDARD; PRT; 754 AA.

AC P46875; Q9FH38;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Kinesin 3 (Kinesin-like protein C).

GN ATK3 OR KATC OR AT5G54670 OR K5F14.1 OR MRB17.18.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.

OC NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;

RX MEDLINE=94355659; PubMed=8075402;

RA Mitsui H., Nakatani K., Yamaguchi-Shinozaki K., Shinozaki K.,

RA Nishikawa K., Takahashi H.,

RT "Sequencing and characterization of the kinesin-related genes katB

RT and katC of Arabidopsis thaliana."

RL Plant Mol. Biol. 25:865-876(1994).

RN [2]

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 16:54:27 ; Search time 4.66667 Seconds
(without alignments)
90.694 Million cell updates/sec

Title: US-10-014-658-1
Perfect score: 41
Sequence: 1 STALEAIGR 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	90.2	535	1 VHTJ_BPT3	P20323 bacteriophage
2	37	90.2	536	1 VHTJ_BPT7	P03728 bacteriophage
3	32	78.0	334	1 YE96_MYCTU	P71777 mycobacteri
4	32	78.0	575	1 TERM_ADEG1	Q64752 avian adeno
5	32	78.0	745	1 ATK2_ARATH	P46864 arabidopsis
6	32	78.0	754	1 ATK3_ARATH	P46875 arabidopsis
7	32	78.0	1313	1 MIPI_SCHPO	P87141 schizosacch
8	31	75.6	322	1 K6P1_THETH	P21777 thermus the
9	31	75.6	547	1 COCH_CHICK	O42163 gallus gall
10	31	75.6	2335	1 TORI_SCHPO	O14356 schizosacch
11	30	73.2	95	1 DEF3_RABIT	P01376 oryctolagus
12	30	73.2	95	1 DEF4_RABIT	P01377 oryctolagus
13	30	73.2	249	1 TPIS_TREPA	O83548 treponema p
14	30	73.2	327	1 PEN_HALN1	Q9hq27 halobacteri
15	30	73.2	340	1 GBBI_DROME	P26308 drosophila
16	30	73.2	614	1 FEOR_SYNY3	P73182 synchocyst
17	30	73.2	736	1 DHB4_HUMAN	P51659 homo sapien
18	30	73.2	758	1 YP58_YEAST	Q99299 saccharomyc
19	30	73.2	944	1 Y166_UREPA	Q99qx7 ureaplasma
20	29	70.7	81	1 ATPL_CLOAB	O08310 clostridium
21	29	70.7	207	1 HIS5_AQUAE	O66943 aquifex aeo
22	29	70.7	217	1 CLD9_MOUSE	Q92087 mus musculu
23	29	70.7	250	1 REPA_AGRTU	P15394 agrobacteri
24	29	70.7	340	1 APBE_BUCAP	O85292 buchnera ap
25	29	70.7	378	1 YN91_ANASP	P46080 anabaena sp
26	29	70.7	387	1 YMBI_CABEL	Q03599 caenorhabdi
27	29	70.7	430	1 ENO_XANAC	Q8p1s0 xanthomonas
28	29	70.7	430	1 ENO_XANCP	Q8p923 xanthomonas
29	29	70.7	430	1 HISX_BRUSU	Q8g2r2 brucella su
30	29	70.7	430	1 HIXI_RHILO	Q989e7 rhizobium l
31	29	70.7	430	1 MURA_RHIME	Q92827 rhizobium m
32	29	70.7	432	1 HISX_AGRTS	Q8uhx1 agrobacteri
33	29	70.7	433	1 HISX_METAC	Q8t141 methanosarc

P97043 leptospira
Q8ru00 brevibacill
Q11157 mycobacteri
P47825 drosophila
Q9hwn1 halobacteri
P20880 human immun
P42352 synchocyst
Q8dhw5 synchococc
Q8x034 neurospora
P31485 yersinia en
P38852 saccharomyc
Q06861 mycobacteri

ALIGNMENTS

```
RESULT 1
VHTJ_BPT3
ID VHTJ_BPT3 STANDARD; PRT; 535 AA.
AC P20323;1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-FEB-1991 (Rel. 17, Last annotation update)
DE Head-TO-tail joining protein.
GN 8.
OS Bacteriophage T3.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;
OC T7-like viruses.
OX NCBI_TaxID=10759;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Luria;
RX MEDLINE=90133923; PubMed=2614843;
RA Beck P.J., Gonzalez S., Ward C.L., Molineux I.J.;
RT "Sequence of bacteriophage T3 DNA from gene 2.5 through gene 9.";
RL J. Mol. Biol. 210:887-701(1989).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X17255; CAA35152.1; -.
DR PIR; S07521; S07521.
KW Structural protein.
SQ SEQUENCE 535 AA; 58620 MW; 1791F9551535535C CRC64;

Query Match 90.2%; Score 37; DB 1; Length 535;
Best Local Similarity 88.9%; Pred. No. 2.4;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 STALEAIGR 9
DB 427 STGLEAIGR 435

RESULT 2
VHTJ_BPT7
ID VHTJ_BPT7 STANDARD; PRT; 536 AA.
AC P03728;1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 21-JUL-1986 (Rel. 01, Last annotation update)
DE Head-TO-tail joining protein.
GN 8.
OS Bacteriophage T7.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;
OC T7-like viruses.
OX NCBI_TaxID=10760;
```

```

RX MEDLINE=21481970; PubMed=11598067;
RA Wang L., Qu W., Reeves P.R.;
RT "Sequence Analysis of Four Shigella boydii O-Antigen Loci: Implication
RT for Escherichia coli and Shigella Relationships.";
RL Infect Immun 69:6923-6930(2001).
DR EMBL; AF402313; AAL27326.1; -.
DR InterPro; IPR001173; Glyco_transf_2; 1.
DR Pfam; PF00535; Glycos_transf_2; 1.
KW Transference.
SQ SEQUENCE 253 AA; 28962 MW; B7EA13F4C6383766 CRC64;

Query Match 78.0%; Score 32; DB 2; Length 253;
Best Local Similarity 77.8%; Pred. No. 1.1e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 STALEAIGR 9
Db 186 ATALEFIQR 194

RESULT 13
Q92UW3 PRELIMINARY; PRT; 261 AA.
AC Q92UW3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative transcriptional regulator, gntR family protein.
GN RE0973 OR SMB21533.
OS Rhizobium meliloti (Sinorhizobium meliloti).
CG Plasmid pSymb (megaplasmid 2).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396508; PubMed=11481431;
RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
RA Vorholter F.J., Hernandez-Lucas I., Becker A., Gouzy J.,
RA Golding B., Puchler A.,
RT "The complete sequence of the 1.693-kb pSymb megaplasmid from the N2-
RT fixing endosymbiont Sinorhizobium meliloti.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
DR EMBL; AL603645; CAC49373.1; -.
DR InterPro; IPR001034; HTH_Deor.
DR Pfam; PF00455; deor; 1.
DR PRINTS; PR00037; HTHLACR.
DR SMART; SM00420; HTH_Deor; 1.
DR PROSITE; PS00894; HTH_Deor_Family; 1.
KW Plasmid; Hypothetical protein; Complete proteome.
SQ SEQUENCE 261 AA; 27626 MW; F893FF9A2DD773DB CRC64;

Query Match 78.0%; Score 32; DB 16; Length 261;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 ALEAIGR 9
Db 163 ALEAIGR 169

RESULT 14
Q9F641 PRELIMINARY; PRT; 270 AA.
AC Q9F641;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE MxGB.
GN MxGB.
OC Stigmatella aurantiaca.
OC Bacteria; Proteobacteria; Deltaproteobacteria; Myxococcales;

Cystobacterineae; Cystobacteraceae; Stigmatella.
OC NCBI_TaxID=41;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SG a15;
RX MEDLINE=20485524; PubMed=11029592;
RA Slakowski B., Kunze B., Nordsiek G., Blocker H., Hofle G., Muller R.;
RT "The myxochelin iron transport regulon of the myxobacterium
RT Stigmatella aurantiaca Sg a15.";
RL Eur. J. Biochem. 267:6476-6485(2000).
DR EMBL; AF299336; AAG31125.1; -.
DR InterPro; IPR007037; SIP.
DR Pfam; PF04954; SIP; 1.
SQ SEQUENCE 270 AA; 29960 MW; 617295D514178F0D CRC64;

Query Match 78.0%; Score 32; DB 2; Length 270;
Best Local Similarity 77.8%; Pred. No. 1.2e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 STALEAIGR 9
Db 151 ATALPAIGR 159

RESULT 15
Q8YZA8 PRELIMINARY; PRT; 293 AA.
AC Q8YZA8;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Cyanophycinase.
GN ALLO571.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL; AP003582; BAB72529.1; -.
DR InterPro; IPR005320; Peptidase_S51.
DR Pfam; PF03575; Peptidase_S51; 1.
KW Complete proteome.
SQ SEQUENCE 293 AA; 31257 MW; F1968EFAE0BCA732 CRC64;

Query Match 78.0%; Score 32; DB 16; Length 293;
Best Local Similarity 77.8%; Pred. No. 1.3e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 STALEAIGR 9
Db 88 STALEAIGR 96

Search completed: December 11, 2003, 18:27:50
Job time : 27.3333 secs

```

DE Agouti-related protein (Fragment).
OS Coturnix coturnix japonica (Japanese quail).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Coturnix.
OX NCBI_TaxID=93934;
RN [1] _
RP SEQUENCE FROM N.A.
RC TISSUE-Hypothalamus;
RA Boswell T., Li Q., Takeuchi S.;
RT "Neurons expressing neuropeptide Y mRNA in the infundibular
RT hypothalamus of Japanese quail are activated by fasting and co-express
RT agouti-related protein mRNA."
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY048849; AAL06600.1; -
FT NON TER 1
SQ SEQUENCE 71 AA; 7866 MW; BEF043CE5A71B22 CRC64;
Query Match 80.5%; Score 33; DB 13; Length 71;
Best Local Similarity 77.8%; Pred. No. 18;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 STALEAIGR 9
Db 10 STALQAAGR 18
RESULT 6
Q9PWG2 PRELIMINARY; PRT; 154 AA.
AC Q9PWG2
DT 01-MAY-2000 (TREMELrel. 13, Created)
DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)
DT 01-OCT-2002 (TREMELrel. 22, Last annotation update)
DE Agouti-related protein.
GN AGRP.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1] _
RP SEQUENCE FROM N.A.
RC STRAIN=White Leghorn;
RX MEDLINE=20335182; PubMed=10771094;
RA Takeuchi S., Teshigawara K., Takahashi S.;
RT "Widespread expression of Agouti-related protein (AGRP) in the
RT chicken: a possible involvement of AGRP in regulating peripheral
RT melanocortin systems in the chicken."
RL Biochim. Biophys. Acta 1496:261-269(2000).
DR EMBL: AB029443; BAA82257.1; -
SQ SEQUENCE 154 AA; 16744 MW; F810B99724120FDD CRC64;
Query Match 80.5%; Score 33; DB 13; Length 154;
Best Local Similarity 77.8%; Pred. No. 40;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 STALEAIGR 9
Db 93 STALQAAGR 101
RESULT 7
Q9W7R0 PRELIMINARY; PRT; 165 AA.
AC Q9W7R0
DT 01-NOV-1999 (TREMELrel. 12, Created)
DT 01-NOV-1999 (TREMELrel. 12, Last sequence update)
DT 01-OCT-2002 (TREMELrel. 22, Last annotation update)
DE Agouti-related protein.
GN AGRP.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1] _
RP SEQUENCE FROM N.A.
RC STRAIN=White Leghorn;
RX MEDLINE=20335182; PubMed=10771094;
RA Takeuchi S., Teshigawara K., Takahashi S.;
RT "Widespread expression of Agouti-related protein (AGRP) in the
RT chicken: a possible involvement of AGRP in regulating peripheral
RT melanocortin systems in the chicken."
RL Biochim. Biophys. Acta 1496:261-269(2000).
DR EMBL: AB029443; BAA82256.1; -
SQ SEQUENCE 165 AA; 17905 MW; CD2B082F5651ADD7 CRC64;
Query Match 80.5%; Score 33; DB 13; Length 165;
Best Local Similarity 77.8%; Pred. No. 43;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 STALEAIGR 9
Db 104 STALQAAGR 112
RESULT 8
Q9E6P9 PRELIMINARY; PRT; 319 AA.
AC Q9E6P9
DT 01-MAR-2001 (TREMELrel. 16, Created)
DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)
DT 01-MAR-2002 (TREMELrel. 20, Last annotation update)
DE U18 capsid protein-like protein.
GN MV030 OR U18.
OS Turkey herpesvirus.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Marek's disease-like viruses.
OX NCBI_TaxID=10390;
RN [1] _
RP SEQUENCE FROM N.A.
RC STRAIN=Md5;
RX MEDLINE=20392152; PubMed=10933706;
RA Tulman E.R., Afonso C.L., Lu Z., Zaak L., Rock D.L., Kutish G.F.;
RT "The genome of a very virulent Marek's disease virus."
RT J. Virol. 74:7980-7988(2000).
RN [2] _
RP SEQUENCE FROM N.A.
RC STRAIN=Md5;
RA Tulman E.R., Afonso C.L., Lu Z., Zaak L., Rock D.L., Kutish G.F.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3] _
RP SEQUENCE FROM N.A.
RC STRAIN=RB1B;
RA Kut E., Rasschaert D.;
RT "Capsid of Marek Disease Virus (MDV): assembling in baculovirus system
RT and determination an immunodominant region."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF243438; AAL14210.1; -
DR EMBL: AF439271; AAL37971.1; -
DR InterPro; IPR002690; Herpes V23.
DR Pfam; PF01802; Herpes V23; I.
SQ SEQUENCE 319 AA; 34664 MW; 64F9AD7D92DB38B CRC64;
Query Match 80.5%; Score 33; DB 12; Length 319;
Best Local Similarity 77.8%; Pred. No. 87;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 STALEAIGR 9
Db 157 STALEAIGR 165
RESULT 9
Q91BV7

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 18:03:30 ; Search time 20.3333 Seconds
(without alignments)
114.220 Million cell updates/sec

Title: US-10-014-658-1

Perfect score: 41

Sequence: 1 STALEAIGR 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPTREMBL 23:*
- 1: sp archaea:*
 - 2: sp bacteria:*
 - 3: sp fungi:*
 - 4: sp human:*
 - 5: sp invertebrate:*
 - 6: sp mammal:*
 - 7: sp mhc:*
 - 8: sp organelle:*
 - 9: sp phase:*
 - 10: sp plant:*
 - 11: sp rodent:*
 - 12: sp virus:*
 - 13: sp vertebrate:*
 - 14: sp unclassified:*
 - 15: sp rvirus:*
 - 16: sp bacteriap:*
 - 17: sp archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	37	90.2	535	Q9T110	Q9T110 bacteriophage
2	35	85.4	415	Q92PU9	Q92PU9 rhizobium m
3	35	85.4	540	Q8PQ63	Q8PQ63 xanthomonas
4	34	82.9	307	O68124	O68124 burkholderi
5	33	80.5	71	Q30WY7	Q30WY7 coturnix co
6	33	80.5	154	Q3PFWG2	Q3PFWG2 gallus gall
7	33	80.5	165	Q3W7R0	Q3W7R0 gallus gall
8	33	80.5	319	Q9E6P9	Q9E6P9 turkey herp
9	33	80.5	319	Q9IBV7	Q9IBV7 turkey herp
10	33	80.5	621	Q17492	Q17492 caenorhabdi
11	33	80.5	702	Q26412	Q26412 methanobact
12	32	78.0	253	Q93CT8	Q93CT8 shigella bo
13	32	78.0	261	Q92UW3	Q92UW3 rhizobium m
14	32	78.0	270	Q9F641	Q9F641 stigmatella
15	32	78.0	293	Q9YZA8	Q9YZA8 anabaena sp
16	32	78.0	327	Q9CBM8	Q9CBM8 mycobacteri

17	32	78.0	351	16	Q92TM4	Q92tm4 rhizobium m
18	32	78.0	362	16	Q927R2	Q927r2 listeria in
19	32	78.0	362	16	Q8Y474	Q8y474 listeria mo
20	32	78.0	393	16	Q8K7L1	Q8k7l1 bacillus ba
21	32	78.0	397	16	Q8DMK1	Q8dmk1 synecococc
22	32	78.0	467	16	Q8DDV0	Q8ddv0 vibrio vuln
23	32	78.0	469	16	Q8PK12	Q8pk12 xanthomonas
24	32	78.0	480	16	Q8KNM0	Q8knm0 vibrio chol
25	32	78.0	725	2	Q8KY01	Q8ky01 rhodospseudo
26	32	78.0	896	5	O62348	O62348 caenorhabdi
27	32	78.0	906	5	O45718	O45718 caenorhabdi
28	32	78.0	1804	10	Q9ZSD0	Q9zsd0 lactuca sat
29	32	78.0	1810	10	Q9ZSD1	Q9zsd1 lactuca sat
30	32	78.0	1813	10	Q9ZSC9	Q9zsc9 lactuca sat
31	32	78.0	1847	10	Q9ZT69	Q9zt69 lactuca sat
32	32	78.0	2410	5	Q8SR14	Q8sr14 encephalito
33	32	78.0	2867	5	Q9N2M3	Q9n2m3 plasmodium
34	31	75.6	97	2	Q8GAL7	Q8gal7 arthrobacte
35	31	75.6	177	11	Q9D096	Q9d096 mus musculu
36	31	75.6	199	4	Q9Y3B5	Q9y3b5 homo sapien
37	31	75.6	258	16	Q982W4	Q982w4 rhizobium l
38	31	75.6	259	16	Q8NUA1	Q8nuu1 corynebacte
39	31	75.6	262	16	Q8P4U9	Q8p4u9 xanthomonas
40	31	75.6	264	16	Q8PPX5	Q8ppx5 xanthomonas
41	31	75.6	297	11	Q91V64	Q91v64 mus musculu
42	31	75.6	288	4	Q96CN7	Q96cn7 homo sapien
43	31	75.6	288	16	Q9KIA9	Q9kya9 streptomyce
44	31	75.6	300	16	Q914J4	Q914j4 pseudomonas
45	31	75.6	300	16	Q92N87	Q92n87 rhizobium m

ALIGNMENTS

RESULT 1
Q9T110 PRELIMINARY; PRT; 535 AA.

AC Q9T110; 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE Head-to-tail joining protein.
GN 8.

OS Bacteriophage phiye03-12.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;
OC T7-like viruses.
OX NCBI_TaxID=110457;
RN {}

RP SEQUENCE FROM N.A.
RA Pajunen M.I.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.

RN {}
RP SEQUENCE FROM N.A.
RX MEDLINE=21125554; PubMed=1122590;
RA Pajunen M.I.; Kiljunen S.J.; Soederholm M.E.L.; Skurnik M.;
RT "Complete genomic sequence of the lytic bacteriophage phiye03-12 of
RT Yersinia enterocolitica serotype O:3.";
RL J. Bacteriol. 183:1928-1937(2001).
DR EMBL; AJ251805; CAB63627.1; -.
SQ SEQUENCE 535 AA; 58649 MW; 03E868B8D64B2BB0 CRC64;

Query Match 90.2%; Score 37; DB 9; Length 535;
Best Local Similarity 88.9%; Pred. NO. 20;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 STALEAIGR 9
|||
Db 427 STGLEAIGR 435

RESULT 2

Q92PU9 PRELIMINARY; PRT; 415 AA.
ID Q92PU9

ABG47941
ID ABG47941 standard; Peptide; 54 AA.
XX AC ABG47941;
XX DT 25-FEB-2003 (first entry)
XX DE Human liver peptide, SEQ ID No 26589.
XX DE Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
XX KW hypercholesterolaemia; coronary heart disease.
XX OS Homo sapiens.
XX PN WO200157273-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00664.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-48898/53.
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analysing gene expression in human adult liver -
XX PS Claim 27; SEQ ID No 26589; 659pp; English.
XX CC The invention relates to a single exon nucleic acid probe (SENP) (I) for
XX CC measuring human gene expression in a sample derived from human adult
XX CC liver, comprising one of 13109 defined nucleotide sequences given in the
XX CC specification (or complements/ fragments). The probe hybridises at high
XX CC stringency to a nucleic acid molecule expressed in the human adult
XX CC liver. (I) may be used for predicting, measuring and displaying gene
XX CC expression in samples derived from human adult liver. The genes
XX CC identified may be involved in genetic liver diseases such as cirrhosis,
XX CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which
XX CC is associated with coronary heart disease. ABG47348-ABG5930 represent
XX CC human liver single exon encoded peptides of the invention.
XX CC Note: The sequence information for this patent does not appear in the
XX CC printed specification but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 54 AA;
Query Match 76.2%; Score 32; DB 22; Length 54;
Best Local Similarity 55.6%; Pred. No. 13;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
QY 1 STEVEAAGR 9
Db 30 NTEIQSAGR 38
RESULT 13
ABB27920
ID ABB27920 standard; Peptide; 54 AA.
XX AC ABB27920;
XX DT 01-FEB-2002 (first entry)
XX DE Peptide #598 encoded by human foetal liver single exon probe.
DE Human peptide #571 encoded by breast cell single exon nucleic acid probe.
XX Human; microarray; single exon probe; gene expression; breast;
XX disease; cancer.
XX OS Homo sapiens.
XX PN WO200157271-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00662.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-496933/54.
XX PT New spatially-addressable set of single exon nucleic acid probes,
XX PT useful for measuring gene expression in sample derived from human
XX PT breast, comprises number of single exon nucleic acid probes -
XX PS Claim 27; SEQ ID NO 10889; 327pp + sequence listing; English.
XX CC The invention relates to a spatially-addressable set of single exon
XX CC nucleic acid probes for measuring gene expression in a sample derived
XX CC from human breast and BT 474 cells. The method involves contacting
XX CC the probes with a collection of detectably labelled nucleic acids
XX CC derived from mRNA of human breast, and then measuring the label
XX CC bound to each probe of the microarray. The probes are useful for
XX CC verifying the expression of regions of genomic DNA predicted to
XX CC encode proteins, they are useful for gene discovery, and for
XX CC determining predigestion and/or prognosing breast disease. Gene
XX CC expression analysis is useful for assessing the toxicity of chemical
XX CC agents on cells. The microarray of this invention presents a far greater
XX CC diversity of probes for measuring gene expression, with far less bias
XX CC than expressed sequence tag microarrays. The method is suitable for
XX CC rapid production of functional information from genomic sequence. The
XX CC present sequence is a peptide encoded by a single exon nucleic acid
XX CC probe of the invention.
XX CC Note: The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 54 AA;
Query Match 76.2%; Score 32; DB 22; Length 54;
Best Local Similarity 55.6%; Pred. No. 13;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
QY 1 STEVEAAGR 9
Db 30 NTEIQSAGR 38
RESULT 14
ABB33092
ID ABB33092 standard; Peptide; 54 AA.
XX AC ABB33092;
XX DT 04-FEB-2002 (first entry)
XX DE Peptide #598 encoded by human foetal liver single exon probe.

QY 2 TEVEACGR 9
DB 19 TEVEACGR 26

RESULT 8
AAU53975
ID AAU53975 standard; Protein; 314 AA.
XX
AC AAU53975;
XX
DT 27-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein #14871.
XX
KW SAPHO syndrome; synovitis; acne; pustulosis; hypervitosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
PN WO200181581-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001WO-US12865.
XX
PR 21-APR-2000; 2000US-199047P.
XX
PR 02-JUN-2000; 2000US-208841P.
XX
PR 07-JUL-2000; 2000US-218747P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR WPI; 2001-616774/71.
XX
DR N-PSDB; AAS59562.
XX

Propionibacterium acnes polypeptides and nucleic acids useful for
vaccinating against and diagnosing infections, especially useful for
treating acne vulgaris -

Example 1; SEQ ID No 15170; 1069pp; English.

Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
polypeptides. The proteins and their associated DNA sequences are used in
the treatment, prevention and diagnosis of medical conditions caused by
P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
pustulosis, hypervitosis and osteomyelitis), uveitis and endophthalmitis.
P. acnes is also involved in infections of bone, joints and the central
nervous system, however it is particularly involved in the inflammatory
lesions associated with acne vulgaris. A method for detecting the
presence or absence of P. acnes in a patient comprises contacting a
sample with a binding agent that binds to the proteins of the invention
and determining the amount of bound protein in the sample. The
polypeptides may be used as antigens in the production of antibodies
specific for P. acnes proteins. These antibodies can be used to
downregulate expression and activity of P. acnes polypeptides and
therefore treat P. acnes infections. The antibodies may also be used as
diagnostic agents for determining P. acnes presence, for example, by
enzyme linked immunosorbent assay (ELISA).

Note: The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.

Sequence 314 AA;
Query Match 78.6%; Score 33; DB 22; Length 314;
Best Local Similarity 77.8%; Pred. No. 63;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 STEVEACGR 9
DB 192 SPEVDACGR 200

RESULT 9
AAG99962
ID AAG99962 standard; Protein; 527 AA.
XX
AC AAG99962;
XX
DT 27-SEP-2001 (first entry)
XX
DE ERA binding domain polypeptide SEQ ID NO 404.
XX
KW ERA binding domain; Escherichia coli; GTPase; antimicrobial;
KW antibacterial; antibiotic; pathogenesis; infection; vaccine;
KW peptide therapy.
XX
OS Escherichia coli.
XX
PN WO200153458-A2.
XX
PD 26-JUL-2001.
XX
PF 17-JAN-2001; 2001WO-US01786.
XX
PR 18-JAN-2000; 2000US-0176870.
XX
PA (SMIK) SMITHKLINE BEECHAM CORP.
PA (SMIK) SMITHKLINE BEECHAM PLC.
XX
PI Lupas AN, Pearce KH;
XX
DR WPI; 2001-476108/51.
XX

New ERA binding domain polypeptides and polynucleotides encoding them,
useful as research reagents and materials for discovery of treatments
and diagnostics for diseases, or for genetic immunisation -

Claim 1; Page 263-265; 279pp; English.

The present invention relates to ERA binding domain polypeptides
(AAG99559-AAG99989 and AAM00010-AAM00021). The era gene in Escherichia
coli codes for an essential GTPase protein able to autophosphorylate at
serine and/or threonine residues. The protein has potential antimicrobial
and antibacterial activity and is useful in screening for antagonists,
agonists and for compounds with antibiotic activity. The proteins are
also useful in determining their role in pathogenesis of infection,
dysfunction and disease and could be used as part of a vaccine and/or
peptide therapy.

Sequence 527 AA;
Query Match 78.6%; Score 33; DB 22; Length 527;
Best Local Similarity 87.5%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TEVEACGR 9
DB 107 TEVEACGR 114

RESULT 10
AAV44469
ID AAV44469 standard; peptide; 9 AA.
XX
AC AAV44469;
XX
DT 27-MAR-2000 (first entry)
XX
DE Human antithrombin III variant Bb.A (residues 395-393).

XX 18-NOV-1999. /note= "ATIII.N135A Val at 389 is substituted by Glu"

XX 12-MAY-1999; 99WO-US10549. /note= "ATIII.N135A Ile at 390 is substituted by Gly"

XX 12-MAY-1998; 98US-0085197.

XX 05-MAY-1999; 99US-0085197.

XX (BOCK/) BOCK S C.

XX (PICA/) PICARD V.

XX (ZEN/) ZENDEHROUH P.

XX Bock SC, Picard V, Zendeherouh P;

XX WPI; 2000-116274/10.

XX New modified human antithrombin III compounds, used for treating e.g. sepsis, trauma, acute respiratory distress syndrome, restenosis, thrombosis, thromboembolism or stroke -

XX Claim 13; Page 57; 75pp; English.

XX The present sequence is from an antithrombin III (ATIII) variant, SEA derived from human ATIII.N135A cDNA insert of the pBlueBac baculovirus expression construct and comprises residues 385-393. The variant has improved resistance to elastase and IgG-activated neutrophils while retains anti-thrombin and anti-factor Xa activities. It may be expressed as glycoforms with enhanced heparin affinity which target the blood vessel wall more efficiently than ATIIIs with normal heparin affinity. The modified ATIIIs can be used to treat thrombin activation-related pathological symptoms due to sepsis, trauma, acute respiratory distress syndrome, restenosis, thrombosis, thromboembolism and stroke. It can also be used to reduce the risk of reocclusion and restenosis in percutaneous transluminal coronary angioplasty, and thrombosis associated with surgery, ischaemia/reperfusion injury, and coagulation abnormalities in cancer or surgical patients.

XX Sequence 9 AA;

Query Match 85.7%; Score 36; DB 21; Length 9;

Best Local Similarity 88.9%; Pred. No. 9.3e+05;

Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 STEVEAAGR 9

DB 1 STAVEAAGR 9

RESULT 4

AA44475

AC AAY44475 standard; peptide; 9 AA.

XX AAY44475;

XX 27-MAR-2000 (first entry)

XX Human antithrombin III variant 13.C (residues 385-393).

XX Human; antithrombin III; ATIII variant 13.C; elastase-resistant; IgG activated neutrophil resistant; anti-thrombin activity; heparin; anti-factor Xa activity; blood clotting disorder; sepsis; trauma; stroke; thrombin activation-related pathological symptom; restenosis; thrombosis; acute respiratory distress syndrome; thromboembolism; reocclusion.

XX Homo sapiens.

OS Synthetic.

XX Key Location/Qualifiers

XX Key

XX Misc-difference 3

XX /note= "ATIII.N135A Ala at 387 is substituted by Glu"

XX Misc-difference 4

XX /note= "ATIII.N135A Val at 388 is substituted by Leu"

XX Misc-difference 5

FT Misc-difference 6

FT /note= "ATIII.N135A Val at 389 is substituted by Glu"

FT /note= "ATIII.N135A Ile at 390 is substituted by Gly"

XX WO9958098-A2.

XX 18-NOV-1999.

XX 12-MAY-1999; 99WO-US10549.

XX 12-MAY-1998; 98US-0085197.

XX 05-MAY-1999; 99US-0085197.

XX (BOCK/) BOCK S C.

XX (PICA/) PICARD V.

XX (ZEN/) ZENDEHROUH P.

XX Bock SC, Picard V, Zendeherouh P;

XX WPI; 2000-116274/10.

XX New modified human antithrombin III compounds, used for treating e.g. sepsis, trauma, acute respiratory distress syndrome, restenosis, thrombosis, thromboembolism or stroke -

XX Claim 13; Page 57; 75pp; English.

XX The present sequence is from an antithrombin III (ATIII) variant, 13.C derived from human ATIII.N135A cDNA insert of the pBlueBac baculovirus expression construct and comprises residues 385-393. The variant has improved resistance to elastase and IgG-activated neutrophils while retains anti-thrombin and anti-factor Xa activities. It may be expressed as glycoforms with enhanced heparin affinity which target the blood vessel wall more efficiently than ATIIIs with normal heparin affinity. The modified ATIIIs can be used to treat thrombin activation-related pathological symptoms due to sepsis, trauma, acute respiratory distress syndrome, restenosis, thrombosis, thromboembolism and stroke. It can also be used to reduce the risk of reocclusion and restenosis in percutaneous transluminal coronary angioplasty, and thrombosis associated with surgery, ischaemia/reperfusion injury, and coagulation abnormalities in cancer or surgical patients.

XX Sequence 9 AA;

Query Match 83.3%; Score 35; DB 21; Length 9;

Best Local Similarity 77.8%; Pred. No. 9.3e+05;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 STEVEAAGR 9

DB 1 STELEGAGR 9

RESULT 5

AA65085

ID AAB65085 standard; peptide; 42 AA.

XX AAB65085;

XX 23-MAR-2001 (first entry)

XX Gene #19 associated peptide #3.

XX Secreted protein; gene therapy; vaccine; cancer; leukemia; autoimmune disease; allergy; inflammation; graft rejection; hyperproliferation; cardiovascular; infection.

XX Homo sapiens.

OS WO200075375-A1.

XX 14-DEC-2000.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 17:22:25 ; Search time 25 Seconds
(without alignments)
57.142 Million cell updates/sec

Title: US-10-014-658-2

Perfect score: 42

Sequence: 1 STEVEAGR 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_19Jun03.*

- 1: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1980.DAT.*
- 2: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1981.DAT.*
- 3: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1982.DAT.*
- 4: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1983.DAT.*
- 5: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1984.DAT.*
- 6: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1985.DAT.*
- 7: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1986.DAT.*
- 8: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1987.DAT.*
- 9: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1988.DAT.*
- 10: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1989.DAT.*
- 11: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1990.DAT.*
- 12: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1991.DAT.*
- 13: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1992.DAT.*
- 14: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1993.DAT.*
- 15: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1994.DAT.*
- 16: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1995.DAT.*
- 17: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1996.DAT.*
- 18: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1997.DAT.*
- 19: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1998.DAT.*
- 20: /SIDSI/gcgdata/geneseq/geneseq-embl/AA1999.DAT.*
- 21: /SIDSI/gcgdata/geneseq/geneseq-embl/AA2000.DAT.*
- 22: /SIDSI/gcgdata/geneseq/geneseq-embl/AA2001.DAT.*
- 23: /SIDSI/gcgdata/geneseq/geneseq-embl/AA2002.DAT.*
- 24: /SIDSI/gcgdata/geneseq/geneseq-embl/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42	100.0	9	21	AAV44466 Human antithrombin
2	38	90.5	9	21	AAV44474 Human antithrombin
3	36	85.7	9	21	AAV44467 Human antithrombin
4	35	83.3	9	21	AAV44475 Human antithrombin
5	35	83.3	42	22	AAV44475 Gene #19 associate
6	35	83.3	267	22	AAV44475 Gene #19 associate
7	34	81.0	66	22	AAV44475 N. gonorrhoeae ami
8	33	78.6	314	22	AAV44475 Propionibacterium
9	33	78.6	527	22	AAV44475 ERA binding domain

10	32	76.2	9	21	AAV44469 Human antithrombin
11	32	76.2	9	21	AAV44473 Human antithrombin
12	32	76.2	54	22	ABG47941 Human liver peptid
13	32	76.2	54	22	ABG47941 Human liver peptid
14	32	76.2	54	22	ABG47941 Human liver peptid
15	32	76.2	54	22	ABG47941 Human liver peptid
16	32	76.2	54	22	ABG47941 Human liver peptid
17	32	76.2	54	22	ABG47941 Human liver peptid
18	32	76.2	54	22	ABG47941 Human liver peptid
19	32	76.2	54	22	ABG47941 Human liver peptid
20	32	76.2	54	22	ABG47941 Human liver peptid
21	32	76.2	54	22	ABG47941 Human liver peptid
22	32	76.2	54	22	ABG47941 Human liver peptid
23	32	76.2	54	22	ABG47941 Human liver peptid
24	32	76.2	54	22	ABG47941 Human liver peptid
25	32	76.2	54	22	ABG47941 Human liver peptid
26	32	76.2	54	22	ABG47941 Human liver peptid
27	32	76.2	54	22	ABG47941 Human liver peptid
28	32	76.2	54	22	ABG47941 Human liver peptid
29	32	76.2	54	22	ABG47941 Human liver peptid
30	32	76.2	54	22	ABG47941 Human liver peptid
31	31	73.8	104	23	ABG59996 Arabidopsis cell c
32	31	73.8	167	24	ABP57965 Horse liver apofe
33	31	73.8	199	23	ABP51378 Human MDP1 SEQ ID
34	31	73.8	290	23	ABP27005 Streptococcus poly
35	31	73.8	317	16	ABP28510 Bovine herpes viru
36	31	73.8	339	23	ABP28510 Streptococcus poly
37	31	73.8	380	15	ABP48063 Sequence of polype
38	31	73.8	412	15	ABP53966 P-selectin ligand.
39	31	73.8	416	23	ABP55981 Bifidobacterium lo
40	30	71.4	43	22	ABP55854 Human testicular a
41	30	71.4	43	22	ABP55152 Human reproductive
42	30	71.4	80	21	AAV44031 Arabidopsis thalia
43	30	71.4	139	22	ABG25462 Novel human diagn
44	30	71.4	164	22	AAV441505 Human immune/haema
45	30	71.4	213	23	ABV47813 Listeria monocytog

ALIGNMENTS

RESULT 1

AAV44466

ID AAV44466 standard; peptide; 9 AA.

XX AAV44466;

AC AAV44466;

XX AAV44466;

XX AAV44466;

XX AAV44466;

XX AAV44466;

XX AAV44466;

XX AAV44466;

XX AAV44466;

XX AAV44466;

XX AAV44466;

XX AAV44466;

XX AAV44466;

XX AAV44466;

XX AAV44466;

XX AAV44466;

XX AAV44466;

XX AAV44466;

XX AAV44466;

XX AAV44466;

XX AAV44466;

XX AAV44466;

XX AAV44466;

XX AAV44466;

XX AAV44466;

XX AAV44466;

XX AAV44466;

XX AAV44466;

XX AAV44466;

XX AAV44466;

XX AAV44466;

XX AAV44466;

XX AAV44466;

XX AAV44466;

XX AAV44466;

XX AAV44466;

XX AAV44466;

XX AAV44466;

ADDRESS: Hamilton, Brook, Smith and Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/421,754
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/304,081
FILING DATE: 01-SEP-1994
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/089,755
FILING DATE: 09-JUL-1993
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/304,081
FILING DATE: 01-SEP-1994
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/911,845
FILING DATE: 10-JUL-1992
ATTORNEY/AGENT INFORMATION:
NAME: Brook, David E
REGISTRATION NUMBER: 22,592
REFERENCE/DOCKET NUMBER: EBC92-03AZX
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 453 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-421-754-2

Query Match 71.4%; Score 30; DB 1; Length 453;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 EVERAGR 9
Db 265 EVDAGR 271

RESULT 13
US-08-421-791-2
Sequence 2, Application US/08421791
Patent No. 5879914
GENERAL INFORMATION:
APPLICANT: Rambosek, John
APPLICANT: Piddington, Chris
APPLICANT: Kovacevich, Brian R
APPLICANT: Young, Kevin D
APPLICANT: Denome, Sylvia A
TITLE OF INVENTION: Recombinant DNA Encoding A
TITLE OF INVENTION: Desulfurization Biocatalyst
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESS: Hamilton, Brook, Smith and Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/421,791
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/304,081
FILING DATE: 01-SEP-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/089,755
FILING DATE: 09-JUL-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/911,845
FILING DATE: 10-JUL-1992
ATTORNEY/AGENT INFORMATION:
NAME: Brook, David E
REGISTRATION NUMBER: 22,592
REFERENCE/DOCKET NUMBER: EBC92-03AZ2
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 453 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-421-791-2
Query Match 71.4%; Score 30; DB 2; Length 453;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 EVERAGR 9
Db 265 EVDAGR 271

RESULT 14
US-08-851-098-2
Sequence 2, Application US/08851088
Patent No. 5952208
GENERAL INFORMATION:
APPLICANT: Darzius, Aldis
APPLICANT: Xi, Lei
APPLICANT: Childs, John D.
APPLICANT: Monticello, Daniel J.
APPLICANT: Squires, Charles H.
TITLE OF INVENTION: DSZ Gene Expression In Pseudomonas Hosts
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/851,088
FILING DATE: 05-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/835,185
FILING DATE: 07-APR-1997
ATTORNEY/AGENT INFORMATION:

APPLICANT: Keil, G nther
TITLE OF INVENTION: Recombinant Bovine Herpesvirus
FILE REFERENCE: vaccine
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Organon Teknika Corporation
STREET: 1330-A Piccard Drive
CITY: Rockville
STATE: Maryland
COUNTRY: U.S.A.
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
FILING DATE: US/08/344,833
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Blackstone, William B.
REGISTRATION NUMBER: 29.772
TELEPHONE: (301) 258-5200
INFORMATION FOR SEQ ID NO: 2:
LENGTH: 317 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-344-833-2

Query Match 73.8%; Score 31; DB 2; Length 317;
Best Local Similarity 77.8%; Pred. No. 50;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 STEVEAAGR 9
Db 239 TTEVEAAAT 247

RESULT 6
US-08-924-345-2
Sequence 2, Application US/08924345
Patent No. 6224878
GENERAL INFORMATION:
APPLICANT: LEUNG-TACK Patricia
APPLICANT: LEGASTELOIS Isabelle, Christine, Marie-Andree
APPLICANT: AUDONNET Jean-Christophe, Francis
APPLICANT: RIVIERE Michel, Emile, Albert
TITLE OF INVENTION: Mutants and vaccines of the Infectious
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: LARSON AND TAYLOR
STREET: 727 SOUTH TWENTY-THIRD STREET
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
FILING DATE: US/08/924,345
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/199,172

FILING DATE: 09-AUG-1994
APPLICATION NUMBER: FR 92 07930
FILING DATE: 26-JUN-1992
ATTORNEY/AGENT INFORMATION:
NAME: SARRO, THOMAS P.
REGISTRATION NUMBER: 19396
REFERENCE/DOCKET NUMBER: XI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 920-7200
TELEFAX: (703) 892-8428
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 380 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-924-345-2

Query Match 73.8%; Score 31; DB 3; Length 380;
Best Local Similarity 77.8%; Pred. No. 61;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 STEVEAAGR 9
Db 237 TTEVEAAAT 245

RESULT 7
US-09-252-991A-32512
Sequence 32512, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 32512
LENGTH: 266
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32512

Query Match 71.4%; Score 30; DB 4; Length 266;
Best Local Similarity 75.0%; Pred. No. 67;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 TEVEAAGR 9
Db 255 TEVSAGR 262

RESULT 8
US-09-761-962A-18
Sequence 18, Application US/09761962A
Patent No. 6500927
GENERAL INFORMATION:
APPLICANT: Memorial Sloan-Kettering Cancer Center
TITLE OF INVENTION: MULTIPLE SPLICE VARIANTS OF THE MU-OFIOLD RECEPTOR GENE
FILE REFERENCE: 830002-2000.2
CURRENT APPLICATION NUMBER: US/09/761,962A
CURRENT FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: 09/743,872
PRIOR APPLICATION NUMBER: PCT/US99/15974
PRIOR FILING DATE: 1999-07-15
NUMBER OF SEQ ID NOS: 46
SOFTWARE: PatentIn version 3.0

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 18:21:01 ; Search time 8.22222 Seconds
(without alignments)
46.313 Million cell updates/sec

Title: US-10-014-658-2

Perfect score: 42

Sequence: 1 STEVEAGR 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310958 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	32	76.2	207	2	US-08-870-518-9
2	32	76.2	459	2	US-08-870-518-1
3	32	76.2	459	2	US-08-870-518-2
4	32	76.2	681	4	US-09-252-991A-18889
5	31	73.8	317	2	US-08-344-833-2
6	31	73.8	380	3	US-08-924-345-2
7	30	71.4	266	4	US-09-252-991A-32512
8	30	71.4	359	4	US-09-761-962A-18
9	30	71.4	399	4	US-09-761-962A-21
10	30	71.4	421	4	US-09-252-991A-17259
11	30	71.4	453	1	US-08-089-755A-2
12	30	71.4	453	1	US-08-421-754-2
13	30	71.4	453	1	US-08-421-791-2
14	30	71.4	453	2	US-08-851-088-2
15	30	71.4	453	2	US-08-851-088-8
16	30	71.4	453	3	US-08-851-089-2
17	30	71.4	453	3	US-08-851-089-9
18	30	71.4	2366	1	US-08-480-604A-10
19	30	71.4	2366	2	US-08-405-496A-10
20	30	71.4	2366	3	US-08-915-136-10
21	30	71.4	2366	4	US-08-957-310-10
22	30	71.4	2366	4	US-10-011-366-10
23	29	69.0	35	3	US-08-658-136-50
24	29	69.0	291	1	US-09-252-991A-26144
25	29	69.0	425	4	US-09-252-991A-29530
26	29	69.0	438	5	PCT-US92-09325-4
27	29	69.0	442	4	US-08-311-731A-150

28	29	69.0	445	1	US-08-700-359-4
29	29	69.0	445	5	PCT-US92-09325-2
30	29	69.0	1113	4	US-09-252-991A-24385
31	29	69.0	1420	2	US-08-540-804-14
32	29	69.0	1420	2	US-08-218-265-14
33	29	69.0	1420	3	US-08-521-872-14
34	29	69.0	1420	3	US-08-590-399-14
35	29	69.0	1805	4	US-09-004-838-92
36	29	69.0	1817	4	US-09-004-838-125
37	29	69.0	1854	4	US-09-004-838-108
38	29	69.0	1890	4	US-09-004-838-88
39	29	69.0	4302	3	US-08-658-136-5
40	29	69.0	4302	4	US-09-052-469-8
41	29	69.0	4302	4	US-08-422-582-8
42	29	69.0	4303	2	US-08-460-751-2
43	29	69.0	4339	4	US-09-052-469-6
44	29	69.0	4339	4	US-08-422-582-6
45	28	66.7	56	5	PCT-US94-06655-5

ALIGNMENTS

RESULT 1
US-08-870-518-9
; Sequence 9, Application US/08870518
; Patent No. 5925566
; GENERAL INFORMATION:
; APPLICANT: Davis, Roger J.
; APPLICANT: Galcheva-Gargova, Zoya
; TITLE OF INVENTION: NON-ACTIVATED RECEPTOR COMPLEX
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/870,518
; FILING DATE: 06-JUN-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/019,219
; FILING DATE: 06-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, Peter J.
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 04020/102001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 207 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-870-518-9

Query Match 76.2%; Score 32; DB 2; Length 207;
Best Local Similarity 55.6%; Pred. No. 18;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 STEVEAGR 9
:|:|:|:|:|

```

; FILE REFERENCE: 11000.1061U
; CURRENT APPLICATION NUMBER: US/10/289,757
; CURRENT FILING DATE: 2002-11-07
; PRIOR APPLICATION NUMBER: 60/337,703
; PRIOR FILING DATE: 2001-11-07
; NUMBER OF SEQ ID NOS: 218
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 172
; LENGTH: 562
; TYPE: PRT
; ORGANISM: Lolium perenne
US-10-289-757-172

Query Match      76.2%; Score 32; DB 12; Length 562;
Best Local Similarity 66.7%; Pred. No. 2.8e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      1 STEVEAAR 9
Db      355 ATDVEAGGR 363

RESULT 14
US-10-142-838B-2
; Sequence 2, Application US/10142838B
; Publication No. US20030124741A1
; GENERAL INFORMATION:
; APPLICANT: Matsushita Electric Industrial Co., Ltd.
; TITLE OF INVENTION: Recombinant cage-like protein, Method for producing the same,
; TITLE OF INVENTION: Precious metal-recombinant cage-like protein complex, Method for
; TITLE OF INVENTION: producing the same and recombinant DNA
; FILE REFERENCE: Apoferritin DNA PRT
; CURRENT APPLICATION NUMBER: US/10/142,838B
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: JP P2001-142983
; PRIOR FILING DATE: 2001-05-14
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 167
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Recombinant Liver Apoferritin
; FEATURE:
; NAME/KEY: MUTAGEN
; LOCATION: (46)
; FEATURE:
; NAME/KEY: MUTAGEN
; LOCATION: (50)
; FEATURE:
; NAME/KEY: MUTAGEN
; LOCATION: (53)
; FEATURE:
; NAME/KEY: MUTAGEN
; LOCATION: (56)
; FEATURE:
; NAME/KEY: MUTAGEN
; LOCATION: (57)
; FEATURE:
; NAME/KEY: MUTAGEN
; LOCATION: (120)
; FEATURE:
; NAME/KEY: MUTAGEN
; LOCATION: (123)
US-10-142-838B-2

Query Match      73.8%; Score 31; DB 15; Length 167;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 STEVEAA 7
```

```

; FILE REFERENCE: 11000.1061U
; CURRENT APPLICATION NUMBER: US/10/384,496
; CURRENT FILING DATE: 2002-11-07
; PRIOR APPLICATION NUMBER: 60/363,163
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 183
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-384-496-8

Query Match      73.8%; Score 31; DB 12; Length 183;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 STEVEAA 7
Db      10 STEVEAA 16

Search completed: December 11, 2003, 18:38:02
Job time : 50.3333 secs
```

```
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 33858
; LENGTH: 54
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC007707.10
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.3
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 3.8
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5.1
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.7
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.8
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.6
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.7
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.5
; OTHER INFORMATION: SWISSPROT HIT: O75312, EVALUATE 3.00e-26
; OTHER INFORMATION: EST HUMAN HIT: BE385905.1, EVALUATE 3.00e-25
; US-09-864-761-33858

Query Match 76.2%; Score 32; DB 9; Length 54;
Best Local Similarity 55.6%; Pred. No. 23;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 STEVEAAGR 9
Db 30 NTEIQSAGR 38

RESULT 7
US-09-988-915-6
; Sequence 6, Application US/09988915
; Patent No. US20020102614A1
; GENERAL INFORMATION:
; APPLICANT: Davis, Roger J.
; APPLICANT: Gangwani, Laxman
; TITLE OF INVENTION: USE OF ZP1 AS A MOLECULAR PROBE FOR SPINAL MUSCULAR ATROPHY
; FILE REFERENCE: 07917-132001
; CURRENT APPLICATION NUMBER: US/09/988,915
; CURRENT FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: US 60/249,745
```

```
; PRIOR FILING DATE: 2000-11-17
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 230
; TYPE: PRT
; ORGANISM: Homo sapiens ZP1 NH2-terminal domain
; US-09-988-915-6

Query Match 76.2%; Score 32; DB 10; Length 230;
Best Local Similarity 55.6%; Pred. No. 1.1e+02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 STEVEAAGR 9
Db 87 NTEIQSAGR 95

RESULT 8
US-10-156-761-12388
; Sequence 12388, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 12388
; LENGTH: 289
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
; US-10-156-761-12388

Query Match 76.2%; Score 32; DB 15; Length 289;
Best Local Similarity 77.8%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 STEVEAAGR 9
Db 235 STAVEVAGR 243

RESULT 9
US-09-925-300-1260
; Sequence 1260, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1260
; LENGTH: 296
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 18:21:27 ; Search time 49.3333 Seconds
(without alignments)
33.929 Million cell updates/sec

Title: US-10-014-658-2

Perfect score: 42

Sequence: 1 STEVEAAGR 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 684280 seqs, 185983659 residues

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

```
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	35	83.3	42	9	US-09-726-643-144
2	35	83.3	42	14	US-10-042-141-144
3	35	83.3	267	9	US-09-726-643-143
4	35	83.3	267	14	US-10-042-141-143
5	34	81.0	395	15	US-10-156-761-11994
6	32	76.2	54	9	US-09-864-761-33858
7	32	76.2	230	10	US-09-988-915-6
8	32	76.2	289	15	US-10-156-761-12388
9	32	76.2	296	10	US-09-925-300-1260
10	32	76.2	390	12	US-10-289-757-80
11	32	76.2	459	10	US-09-988-915-1
12	32	76.2	459	12	US-10-289-757-111
13	32	76.2	562	12	US-10-289-757-172
14	31	73.8	167	15	US-10-142-838B-2
15	31	73.8	183	12	US-10-384-496-8

```
16 73.8 183 12 US-10-384-496-10
17 73.8 183 12 US-10-384-496-14
18 73.8 338 15 US-10-156-761-9235
19 71.4 43 11 US-09-764-891-3810
20 71.4 113 12 US-10-230-331-33
21 71.4 236 10 US-09-738-626-4034
22 71.4 264 10 US-09-738-626-6002
23 71.4 359 9 US-09-761-982-18
24 71.4 359 15 US-10-283-300-18
25 71.4 368 12 US-10-284-400-4
26 71.4 378 9 US-09-815-242-13402
27 71.4 397 12 US-10-284-400-12
28 71.4 399 9 US-09-761-962-21
29 71.4 399 15 US-10-283-300-21
30 71.4 453 12 US-10-286-326-12
31 71.4 453 12 US-10-286-326-12
32 71.4 498 9 US-09-815-242-10556
33 71.4 500 9 US-09-815-242-5007
34 71.4 2366 12 US-10-354-774-10
35 71.4 2366 15 US-10-011-366-10
36 69.0 77 9 US-09-864-761-35739
37 69.0 239 12 US-10-094-749-1700
38 69.0 311 12 US-10-200-562-93
39 69.0 311 12 US-10-237-551-93
40 69.0 311 15 US-10-121-988-93
41 69.0 351 12 US-10-029-386-32088
42 69.0 441 10 US-09-712-363-195
43 69.0 445 10 US-09-738-626-6952
44 69.0 456 9 US-09-815-242-4991
45 69.0 456 9 US-09-815-242-10636
```

ALIGNMENTS

```
RESULT 1
US-09-726-643-144
; Sequence 144, Application US/09726643
; Patent No. US200202849A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 26 Human secreted proteins
; FILE REFERENCE: P204091
; CURRENT APPLICATION NUMBER: US/09/726,643
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: PCT/US00/15187
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: 60/137,725
; PRIOR FILING DATE: 1999-06-07
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 144
; LENGTH: 42
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-726-643-144
```

```
Query Match 83.3%; Score 35; DB 9; Length 42;
Best Local Similarity 77.8%; Pred. No. 4.4;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

QY 1 STEVEAAGR 9

Db 30 SSEVDAGR 38

RESULT 2

```
US-10-042-141-144
; Sequence 144, Application US/10042141
; Publication No. US20020193503A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 26 Human secreted proteins
```

A;Cross-references: EMBL:U11938; NID:q533553; PIDN:AAA99594.1; PID:g1235808
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1994
C;Superfamily: M5 protein

Query Match 73.8%; Score 31; DB 2; Length 101;
Best Local Similarity 66.7%; Pred. No. 17;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 STEVEAQR 9
|:|||||:
Db 30 ATEVKAACQ 38

RESULT 8
PROL

ferritin light chain - horse
C;Species: Equus caballus (domestic horse)
C;Date: 01-Sep-1981 #sequence_revision 08-May-1998 #text_change 16-Jun-2000
C;Accession: S36118; A03267
R;Takeda, S.; Ohta, M.; Ebina, S.; Nagayama, K.
Biochim. Biophys. Acta 1174, 218-220, 1993
A;Title: Cloning, expression and characterization of horse L-ferritin in Escherichia coli
A;Reference number: S36118; MUID:93363645; PMID:8357841
A;Accession: S36118
A;Molecule type: mRNA
A;Residues: 1-175 <TAK>
A;Cross-references: GB:D4523; NID:g406208; PIDN:BAA03396.1; PID:g406209
A;Experimental source: liver
R;Heusterspreute, M.; Crichton, R.R.
FEBS Lett. 129, 322-327, 1981
A;Title: Amino acid sequence of horse spleen apoferritin.
A;Reference number: A91294; MUID:82027739; PMID:7026284
A;Accession: A03267
A;Molecule type: protein
A;Residues: 2-93, 'L', 95-175 <HEU>
A;Experimental source: spleen
R;Clegg, G.A.; Stansfield, R.F.D.; Bourne, P.E.; Harrison, P.M.
Nature 288, 298-300, 1980
A;Title: Helix packing and subunit conformation in horse spleen apoferritin.
A;Reference number: A95239; MUID:81052459; PMID:7432529
A;Contents: annotation; X-ray crystallography, 2.8 angstroms
C;Comment: Ferritin is an intracellular molecule that stores iron in a soluble, nontoxic
tains a central cavity in which the polymeric ferric iron core is deposited.
C;Comment: There are two types of ferritin subunits: L (light) chain and H (heavy) chain.
C;Comment: In horse spleen the light chain is the major chain.
C;Superfamily: ferritin
C;Keywords: acetylated amino end; iron; iron storage; liver; metalloprotein; multimer; s

F;2-175/Product: ferritin light chain #status experimental <Mar>
F;2/Modified site: acetylated amino end (Ser) (in mature form) #status experimental
F;54,57,58,61,64/Binding site: iron (Glu) #status predicted

Query Match 73.8%; Score 31; DB 1; Length 175;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 STEVEAA 7
|:|||||:
Db 10 STEVEAA 16

RESULT 9
JC7238

ferritin protein light chain - guinea pig
C;Species: Cavia porcellus (guinea pig)
C;Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 21-Jul-2000
C;Accession: JC7238
R;Cheng, Q.; Gonzalez, P.; Zigler Jr., J-S.
Biochem. Biophys. Res. Commun. 270, 349-355, 2000
A;Title: High level of ferritin light chain mRNA in lens.
A;Reference number: JC7238; MUID:20218664; PMID:10753629
A;Accession: JC7238
A;Status: preliminary
A;Molecule type: mRNA

A;Residues: 1-175 <CHE>
A;Cross-references: GB:AR233445; PIDN:AAF36408.1
A;Experimental source: strain 13/N

C;Comment: This protein is an important regulator of oxidative stress and a primary fact
d is responsible for regulating the levels of intracellular iron.
C;Superfamily: ferritin
C;Keywords: iron storage

Query Match 73.8%; Score 31; DB 2; Length 175;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 STEVEAA 7
|:|||||:
Db 10 STEVEAA 16

RESULT 10

ferritin light chain - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 28-Aug-1985 #sequence_revision 01-Dec-1995 #text_change 22-Jun-1999
C;Accession: A29575; A03268
R;Leibold, E.A.; Munro, H.N.
J. Biol. Chem. 262, 7335-7341, 1987
A;Title: Characterization and evolution of the expressed rat ferritin light subunit gene
A;Reference number: A29575; MUID:87222340; PMID:3584116
A;Accession: A29575
A;Molecule type: DNA
A;Residues: 1-183 <LEI>
A;Cross-references: GB:J02741; NID:g204132; PIDN:AAA41155.1; PID:g204133
R;Leibold, E.A.; Aziz, N.; Brown, A.J.P.; Munro, H.N.
J. Biol. Chem. 259, 4327-4334, 1984
A;Title: Conservation in rat liver of light and heavy subunit sequences of mammalian fer
A;Reference number: A92474; MUID:84162134; PMID:6546756
A;Accession: A03268
A;Molecule type: mRNA
A;Residues: 2-97, 'K', 99-120, 'QA', 123-154, 'W', 156-183 <LR2>
A;Cross-references: GB:K01930; NID:g204130; PIDN:AAA41154.1; PID:g204131
A;Note: Initiator Met not shown
C;Comment: This ferritin mRNA was isolated from liver parenchymal cells.
C;Comment: The rat light chain has an octopeptide insertion after residue 158 compared v
C;Comment: Ferritin is an intracellular molecule that stores iron in a soluble, nontoxic
tains a central cavity in which the polymeric ferric iron core is deposited.
C;Comment: There are two types of ferritin subunits: L (light) chain and H (heavy) chain
C;Comment: In rat liver the light chain is the major chain.
C;Superfamily: ferritin
C;Keywords: iron storage; liver; multimer

Query Match 73.8%; Score 31; DB 1; Length 183;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 STEVEAA 7
|:|||||:
Db 10 STEVEAA 16

RESULT 11

B33355
ferritin light chain - mouse
C;Species: Mus musculus (house mouse)
C;Date: 20-Dec-1989 #sequence_revision 20-Dec-1989 #text_change 20-Aug-1999
C;Accession: B33355
R;Beaumont, C.; Dugast, I.; Renaudie, F.; Souroujon, M.; Grandchamp, B.
J. Biol. Chem. 264, 7498-7504, 1989
A;Title: Transcriptional regulation of ferritin H and L subunits in adult erythroid and
ferritin shells
A;Reference number: A33355; MUID:89214195; PMID:2708374
A;Accession: B33355
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-183 <BEA>

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 18:19:54 ; Search time 8.33333 Seconds
(without alignments)
103.862 Million cell updates/sec

Title: US-10-014-658-2

Perfect score: 42

Sequence: 1 STEVEAAG 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 76:*

1: Pir1.*

2: Pir2.*

3: Pir3.*

4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	33	78.6	110	2 AE2448	hypothetical prote
2	33	78.6	349	2 T44856	molybdenum transpo
3	33	78.6	435	2 T00126	hypothetical prote
4	33	78.6	738	2 E87627	hypothetical prote
5	33	78.6	1376	1 VCBED6	major capsid prote
6	32	76.2	445	2 B82954	probable dihydroor
7	31	73.8	101	2 S60792	M protein precurs
8	31	73.8	175	1 FRHOL	ferritin light cha
9	31	73.8	175	1 FC7238	ferritin protein 1
10	31	73.8	183	1 FRSTL	ferritin light cha
11	31	73.8	183	2 B33355	ferritin light cha
12	31	73.8	183	2 I54774	ferritin light cha
13	31	73.8	302	2 AE2640	chemotaxis methyl
14	31	73.8	302	2 C97422	cher homolog (AF04
15	31	73.8	378	2 H87333	HlyD family secret
16	31	73.8	380	2 S35785	glycoprotein i - b
17	31	73.8	386	2 S54858	M protein precurs
18	31	73.8	435	2 AG3159	replication protei
19	31	73.8	475	2 C70966	hypothetical prote
20	31	73.8	477	2 AH0748	probable exported
21	31	73.8	952	2 H84583	hypothetical prote
22	31	73.8	1576	2 AE0249	probable hemolysin
23	31	73.8	2499	1 A30788	mannose 6-phosphat
24	30	71.4	113	2 T02040	acidic ribosomal p
25	30	71.4	185	2 E87663	transcriptional regu
26	30	71.4	200	2 E85047	hypothetical prote
27	30	71.4	207	2 B84384	uroporphyrin-III C
28	30	71.4	213	2 AF1222	Salmonella enteric
29	30	71.4	213	2 AI1575	Salmonella enteric

ALIGNMENTS

RESULT 1

AE2448

hypothetical protein alr5141 [imported] - Nostoc sp. (strain PCC 7120)

C;Species: Nostoc sp. PCC 7120

A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002

C;Accession: AE2448

R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iritiguch

Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, M.

DNA Res. 8, 205-213, 2001

A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium An

A;Reference number: AB1807; MUID:21595285; PMID:11759840

A;Accession: AE2448

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-110 <KUR>

A;Cross-references: GB:BA000019; PIDN:BA076840.1; PID:gi7134279; GSPDB:GN00179

A;Experimental source: strain PCC 7120

C;Genetics:

A;Gene: alr5141

Query Match 78.6%; Score 33; DB 2; Length 110;

Best Local Similarity 87.5%; Pred. No. 6.7;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 STEVEAAG 8

Db 88 STEVETAG 95

RESULT 2

T44856

molybdenum transport ATPase (EC 3.6.1.-) [imported] - Arthrobacter nicotinovorans plasm

C;Species: Arthrobacter nicotinovorans

C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 19-Jul-2002

C;Accession: T44856

R;Brandsch, R.

submitted to the EMBL Data Library, February 1999

A;Description: Molybdate-uptake genes and molybdopterin-biosynthesis genes on a bacteri.

A;Reference number: Z22860

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-349 <BRA>

A;Cross-references: EMBL:Y10817; PIDN:CAA71778.1

C;Genetics:

A;Gene: modC

A;Genome: plasmid pAO1

C;Superfamily: molybdenum transport protein modC; ATP-binding cassette homology

C;Keywords: hydrolase; molybdenum transport

Query Match 78.6%; Score 33; DB 2; Length 349;

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; U08850; AAA56671.1; -;
DR EMBL; L37363; AAA99482.1; -;
KW Oxidoreductase, Monooxygenase; Flavoprotein; FMN; Plasmid.
SQ SEQUENCE 453 AA; 49634 MW; 3A45ED097259C82 CRC64;

Query Match 71.4%; Score 30; DB 1; Length 453;
Best Local Similarity 85.7%; Pred. No. 83;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 EVEAAGR 9
||:|||||
DB 265 EVDAAGR 271

RESULT 15

GLPK_DEIRA
ID GLPK_DEIRA STANDARD; PRT; 501 AA.
AC Q9RI38;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glycerol kinase (EC 2.7.1.30) (ATP:glycerol 3-phosphotransferase)
DE (Glycerokinase) (GK).
GN GLPK OR DR1928.
OS Deinococcus radiodurans.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H.S., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
radiodurans R1.";
RL Science 286:1571-1577(1999).
CC -!- FUNCTION: Key enzyme in the regulation of glycerol uptake and
CC metabolism.
CC -!- CATALYTIC ACTIVITY: ATP + glycerol = ADP + glycerol 3-phosphate.
CC -!- PATHWAY: Glycerol utilization; first (rate-limiting) step.
CC -!- SIMILARITY: BELONGS TO THE FUCOKINASE / GLUCOKINASE /
CC GLYCEROKINASE / XYLULOKINASE FAMILY.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; A3002031; AAF11475.1; -;
DR PIR; G75337; G75337.
DR HSSP; P08859; 1GLJ.
DR TIGR; DR1928; -;
DR HAMAP; MF_00186; -; 1.
DR InterPro; IPR000577; FGGY_kin.
DR InterPro; IPR005999; Glycerol_kin.
DR Pfam; PF00370; FGGY; 1.

DR Pfam; PF02782; FGGY_C; 1.
DR TIGRFAM; TIGR01311; glycerol_kin; 1.
DR PROSITE; PS00933; FGGY_KINASES_1; 1.
DR PROSITE; PS00445; FGGY_KINASES_2; 1.
KW Glycerol metabolism; Transferase; Kinase; ATP-binding;
KW Complete proteome. 166 ATP (PROBABLE).
FT NP_BIND 154 166
SQ SEQUENCE 501 AA; 55068 MW; CAA779FFD2ALEC1F CRC64;

Query Match 71.4%; Score 30; DB 1; Length 501;
Best Local Similarity 66.7%; Pred. No. 92;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 STEVEAAGR 9
||:|||||
DB 325 STEIEALAR 333

Search completed: December 11, 2003, 18:20:53
Job time : 7.66667 secs

or send an email to license@sib-sib.ch.

EMBL; 223068; CAA80605.1; -
 PIR; S35785; S35785.
 InterPro; IPR002874; Herpes_g1.
 Pfam; PF01688; Herpes_g1; 1.
 Glycoprotein.
 CARBOHYD 67 N-LINKED (GLCNAC... (POTENTIAL).
 SEQUENCE 380 AA; 39910 MW; ABBE1FB9B430D2BD CRC64;

Query Match 73.8%; Score 31; DB 1; Length 380;
 Best Local Similarity 77.8%; Pred. No. 43;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 STEVEAAGR 9
 :|||||
 :|||||

Db 237 TTEVEAATR 245

RESULT 11
 MPRI_BOVIN
 ID MPRI_BOVIN STANDARD; PRT; 2499 AA.
 AC P08169;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DE Cation-independent mannose-6-phosphate receptor precursor (CI Man-6-P
 DE receptor) (CI-MPR) (Insulin-like growth factor II receptor) (300 kDa
 DE mannose 6-phosphate receptor) (MPR 300) (MPR300).
 GN IGF2R OR M6P.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88115411; PubMed=2963004;
 RA Lobel P., Dahms N.M., Kornfeld S.;
 RT "Cloning and sequence analysis of the cation-independent mannose 6-
 RT phosphate receptor.";
 RL J. Biol. Chem. 263:2563-2570 (1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Killian J.K.;
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 1039-2499 FROM N.A.
 RX MEDLINE=87175648; PubMed=2951738;
 RA Lobel P., Dahms N.M., Breitmeyer J., Chirgwin J.M., Kornfeld S.;
 RT "Cloning of the bovine 215-kDa cation-independent mannose 6-phosphate
 RT receptor.";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:2233-2237 (1987).
 CC -!- FUNCTION: TRANSPORT OF PHOSPHORYLATED LYSOSOMAL ENZYMES FROM
 CC THE GOLGI COMPLEX AND THE CELL SURFACE TO LYSOSOMES. LYSOSOMAL
 CC ENZYMES BEARING PHOSPHOMANNOSYL RESIDUES BIND SPECIFICALLY TO
 CC MANNOSE-6-PHOSPHATE RECEPTORS IN THE GOLGI APPARATUS AND THE
 CC RESULTING RECEPTOR-LIGAND COMPLEX IS TRANSPORTED TO AN ACIDIC
 CC PRELYSOSOMAL COMPARTMENT WHERE THE LOW PH MEDIATES THE DISSOCIATION
 CC OF THE COMPLEX. THIS RECEPTOR ALSO BINDS INSULIN GROWTH FACTOR II.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Lysosomal.
 CC -!- DOMAIN: CONTAINS 15 REPEATING UNITS OF APPROXIMATIVELY 147 AA. THE
 CC MOST HIGHLY CONSERVED REGION WITHIN THE REPEAT CONSISTS OF A
 CC STRETCH OF 13 AA THAT CONTAINS CYSTEINES AT BOTH ENDS.
 CC -!- SIMILARITY: Contains 1 fibronectin type II domain.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@sib-sib.ch.

CC DR EMBL; J03527; AAA30455.1; -
 DR EMBL; AF342811; AAL23908.1; -
 DR PIR; A55908; A30788.
 DR HSP; P02751; 2FN2.
 DR InterPro; IPR000479; CIMR.
 DR InterPro; IPR00562; FN_Type_II.
 DR Pfam; PF00878; CIMR; 13.
 DR Pfam; PF00040; fn2; 1.
 DR PRINTS; PR00013; ENTPELI.
 DR ProDom; PD000395; FN_Type_II; 1.
 DR SMART; SM00059; FN2; 1.
 DR PROSITE; PS00023; FIBRONECTIN_2; 1.
 KW Transmembrane; Transport; Glycoprotein; Repeat; Receptor; Lysosome;
 KW Signal.
 FT SIGNAL 1 44
 FT CHAIN 45 2499
 FT CATION-INDEPENDENT MANNOSE-6-PHOSPHATE
 FT RECEPTOR.
 FT LUMENAL (POTENTIAL).
 FT POTENTIAL.
 FT CYTOPLASMIC (POTENTIAL).
 FT 1.
 FT REPEAT 171 327 2.
 FT REPEAT 328 478 3.
 FT REPEAT 479 629 4.
 FT REPEAT 630 771 5.
 FT REPEAT 772 933 6.
 FT REPEAT 934 1089 7.
 FT REPEAT 1090 1229 8.
 FT REPEAT 1230 1373 9.
 FT REPEAT 1374 1518 10.
 FT REPEAT 1519 1658 11.
 FT REPEAT 1659 1807 12.
 FT REPEAT 1808 1999 13.
 FT REPEAT 2000 2137 14.
 FT REPEAT 2138 2290 15.
 FT DOMAIN 120 120
 FT CARBOHYD 409 409
 FT CARBOHYD 444 444
 FT CARBOHYD 552 552
 FT CARBOHYD 590 590
 FT CARBOHYD 635 635
 FT CARBOHYD 755 755
 FT CARBOHYD 879 879
 FT CARBOHYD 959 959
 FT CARBOHYD 1030 1030
 FT CARBOHYD 1173 1173
 FT CARBOHYD 1255 1255
 FT CARBOHYD 1321 1321
 FT CARBOHYD 1665 1665
 FT CARBOHYD 1766 1766
 FT CARBOHYD 1825 1825
 FT CARBOHYD 2094 2094
 FT CARBOHYD 2145 2145
 FT CARBOHYD 2220 2220
 SQ SEQUENCE 2499 AA; 274526 MW; 3C1C9DF2875159D CRC64;

Query Match 73.8%; Score 31; DB 1; Length 2499;
 Best Local Similarity 66.7%; Pred. No. 2.7e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 STEVEAAGR 9
 :|||:|
 :|||:|

Db 398 STQVKVAGR 406

RESULT 12
 RA2B_MAIZE
 ID RA2B_MAIZE STANDARD; PRT; 113 AA.
 AC O24415;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)

OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=87222340; PubMed=3584116;
RA Leibold E.A., Munro H.N.;
RT "Characterization and evolution of the expressed rat ferritin light
subunit gene and its pseudogene family. Conservation of sequences
within noncoding regions of ferritin genes.";
RL J. Biol. Chem. 262:7335-7341(1987).
RN (2)
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=84162134; PubMed=6546756;
RA Leibold E.A., Ariz N., Brown A.J.P., Munro H.N.;
RT "Conservation in rat liver of light and heavy subunit sequences of
mammalian ferritin. Presence of unique octopeptide in the light
subunit.";
RL J. Biol. Chem. 259:4327-4334(1984).
RN (3)
RP SEQUENCE FROM N.A.
RX MEDLINE=92210224; PubMed=1555892;
RA Denis M.G.;
RT "Isolation of cDNA clones corresponding to genes differentially
expressed in two colon-carcinoma cell lines differing by their
tumorigenicity.";
RL Int. J. Cancer 50:930-936(1992).
CC -!- FUNCTION: Ferritin is an intracellular molecule that stores iron
in a soluble, nontoxic, readily available form. The functional
molecule, which is composed of 24 chains, is roughly spherical and
contains a central cavity into which the polymeric ferric iron
core is deposited.
CC -!- TISSUE SPECIFICITY: IN RAT LIVER THE LIGHT CHAIN IS THE MAJOR
CHAIN.
CC -!- DOMAIN: THE RAT LIGHT CHAIN HAS AN OCTOPEPTIDE INSERTION AFTER
RESIDUE 158 COMPARED WITH OTHER LIGHT CHAINS.
CC -!- MISCELLANEOUS: THERE ARE TWO TYPES OF FERRITIN SUBUNITS: L (LIGHT)
CHAIN AND H (HEAVY) CHAIN. THE MAJOR CHAIN CAN BE LIGHT OR HEAVY,
DEPENDENT ON THE SPECIES AND TISSUE TYPE.
CC -!- SIMILARITY: BELONGS TO THE FERRITIN FAMILY.
CC -!- SIMILARITY: Contains 1 ferritin-like diiron domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
DR EMBL; K01930; AAA41154.1; -;
DR EMBL; J02741; AAA41155.1; -;
DR EMBL; L01122; AAA41152.1; -;
DR PIR; A29575; FRRTL.
DR PIR; I54774; I54774.
DR HSP; P02791; 1IER.
DR InterPro; IPR001519; Ferritin.
DR Pfam; PF00210; ferritin; 1.
DR ProDom; PD000971; Ferritin; 1.
DR PROSITE; PS00540; FERRITIN 1; 1.
DR PROSITE; PS00204; FERRITIN 2; 1.
DR PROSITE; PS09005; FERRITIN LIKE; 1.
KW Iron storage; Iron; Metal-binding.
FT INIT MET 0 0
FT DOMAIN 6 155 FERRITIN-LIKE DIIRON.
FT METAL 53 53 IRON (POTENTIAL).
FT METAL 56 56 IRON (POTENTIAL).
FT METAL 57 57 IRON (POTENTIAL).
FT METAL 60 60 IRON (POTENTIAL).
FT METAL 63 63 IRON (POTENTIAL).
FT METAL

FT CONFLICT 97 97 E -> K (IN REF. 2).
FT CONFLICT 120 121 RT -> OA (IN REF. 2).
FT CONFLICT 125 125 L -> F (IN REF. 3).
FT CONFLICT 154 154 V -> W (IN REF. 2).
FT CONFLICT 155 155 Q -> A (IN REF. 3).
SQ SEQUENCE 182 AA; 20674 MW; 3D3DEBCE5088655B CRC64;
Query Match 73.8%; Score 31; DB 1; Length 182;
Best Local Similarity 100.0%; Pred.No.21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 STEVEAA 7
DB 9 STEVEAA 15
RESULT 8
FRL1_MOUSE STANDARD; PRT; 182 AA.
AC F29391; Q8WUQ8;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Ferritin light chain 1 (Ferritin L subunit 1).
GN FTL1 OR FTL-1 OR FTL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=89214195; PubMed=2708374;
RA Beaumont C., Dugast I., Renaudie F., Souroujon M., Grandchamp B.;
RT "Transcriptional regulation of ferritin H and L subunits in adult
erythroid and liver cells from the mouse. Unambiguous identification
of mouse ferritin subunits and in vitro formation of the ferritin
shells.";
RL J. Biol. Chem. 264:7498-7504(1989).
RN (2)
RP SEQUENCE FROM N.A.
RX MEDLINE=95376418; PubMed=7648356;
RA Renaudie F., Boulanger L., Grandchamp B., Beaumont C.;
RT "Cloning, characterization and expression of mouse ferritin L subunit
gene.";
RL C. R. Acad. Sci., III, Sci. Vie 318:431-437(1995).
RN (3)
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
Brownstein M.J., Ussdin T.B., Roshydyki S., Carninci P., Prange C.,
Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaney S.J.,
Rosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyx S.W.,
Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Fahey J., Helton B., Kettman M., Madan A., Rodrigues S., Sanchez A.,
Whiting M., Madan A., Young A.C., Shevchenko V., Bouffard G.G.,
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: Ferritin is an intracellular molecule that stores iron
in a soluble, nontoxic, readily available form. The functional
molecule, which is composed of 24 chains, is roughly spherical and
contains a central cavity into which the polymeric ferric iron

```
DR MM; 603901; --
DR GO; GO:0005737; C:cytoplasm; TAS.
DR GO; GO:0005730; C:nucleolus; TAS.
DR GO; GO:0008283; P:cell proliferation; TAS.
DR GO; GO:0007165; P:signal transduction; TAS.
DR InterPro; IPR004457; Znf_ZPR1.
DR Pfam; PF03367; ZPR1; 2.
DR ProDom; PD005639; Znf_ZPR1; 2.
DR SMART; SM00709; Zpr1; 2.
DR TIGRFAMs; TIGR00310; ZPR1.znf; 2.
KW Nuclear protein; Zinc-finger.
FT ZN_FING 51 83 C4-TYPE.
FT ZN_FING 259 291 C4-TYPE.
SQ SEQUENCE 459 AA; 50925 MW; E3DB820F490F2835 CRC64;

Query Match 76.2%; Score 32; DB 1; Length 459;
Best Local Similarity 55.6%; Pred. No. 32;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 STEVEAGR 9
DB 87 NTEIQSAGR 95

RESULT 4
ZPR1_MOUSE
ID_ZPR1_MOUSE STANDARD; PRT; 459 AA.
AC Q62384;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zinc-finger protein ZPR1 (Zinc finger protein 259).
GN ZNF259 OR ZFP259 OR ZPR1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Galcheva-Gargova Z., Konstantinov K.N., Wu I.-H., Klier F.G.,
RA Barrett T., Davis R.J.;
RT "Binding of zinc finger protein ZPR1 to the epidermal growth
RT factor receptor.";
RL Science 272:1797-1802(1996).
RN [2]
RP SUBCELLULAR LOCATION.
RX MEDLINE=98437195; PubMed=9763455;
RA Galcheva-Gargova Z., Gangwani L., Konstantinov K.N., Mikrut M.,
RA Theroux S.J., Enoch T., Davis R.J.;
RT "The cytoplasmic zinc finger protein ZPR1 accumulates in the nucleolus
RT of proliferating cells.";
RL Mol. Biol. Cell 9:2963-2971(1998).
CC -1- FUNCTION: MAY BE A SIGNALING MOLECULE THAT COMMUNICATES MITOGENIC
CC SIGNALS FROM THE CYTOPLASM TO THE NUCLEUS.
CC -1- SUBUNIT: BINDS TO THE EGF AND PDGF RECEPTORS. BINDS TO THE
CC ELONGATION FACTOR 1-ALPHA.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC; TRANSLOCATE TO THE NUCLEUS
CC AFTER TREATMENT WITH MITOGENS.
CC -1- SIMILARITY: BELONGS TO THE ZPR1 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U41287; AAC52662.1; --
CC MGD; MG1:1330262; Zfp259.
DR GO; GO:0005737; C:cytoplasm; IDA.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0005515; P:protein binding activity; IPI.

GO; GO:0008270; F:zinc ion binding activity; IDA.
DR InterPro; IPR004457; Znf_ZPR1.
DR Pfam; PF03367; ZPR1; 2.
DR SMART; SM00709; Zpr1; 2.
DR TIGRFAMs; TIGR00310; ZPR1.znf; 2.
KW Nuclear protein; Zinc-finger.
FT ZN_FING 51 83 C4-TYPE.
FT ZN_FING 259 291 C4-TYPE.
SQ SEQUENCE 459 AA; 50715 MW; 771D39C7B806044P CRC64;

Query Match 76.2%; Score 32; DB 1; Length 459;
Best Local Similarity 55.6%; Pred. No. 32;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 STEVEAGR 9
DB 87 NTEIQSAGR 95

RESULT 5
FRIL_BOVIN
ID_FRIL_BOVIN STANDARD; PRT; 174 AA.
AC Q46415;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Ferritin light chain (Ferritin L subunit).
GN FTL.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=Spleen.
RX MEDLINE=98129059; PubMed=9467878;
RA Orino K., Eguchi K., Nakayama T., Yamamoto S., Watanabe K.;
RT "Sequencing of cDNA clones that encode bovine ferritin H and L
RT chains.";
RL Comp. Biochem. Physiol. 118B:667-673(1997).
CC -1- FUNCTION: Ferritin is an intracellular molecule that stores iron
CC in a soluble, nontoxic, readily available form. The functional
CC molecule, which is composed of 24 chains, is roughly spherical and
CC contains a central cavity into which the polymeric ferric iron
CC core is deposited.
CC -1- MISCELLANEOUS: THERE ARE TWO TYPES OF FERRITIN SUBUNITS: L (LIGHT)
CC CHAIN AND H (HEAVY) CHAIN. THE MAJOR CHAIN CAN BE LIGHT OR HEAVY,
CC DEPENDING ON THE SPECIES AND TISSUE TYPE.
CC -1- SIMILARITY: BELONGS TO THE FERRITIN FAMILY.
CC -1- SIMILARITY: Contains 1 ferritin-like diiron domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB003094; BAA24819.1; --
CC HSSP; P02791; IDAT.
DR InterPro; IPR001519; Ferritin.
DR Pfam; PF00210; ferritin; 1.
DR ProDom; PD000971; Ferritin; 1.
DR PROSITE; PS00540; FERRITIN_1; 1.
DR PROSITE; PS00204; FERRITIN_2; 1.
DR PROSITE; PS50905; FERRITIN-LIKE; 1.
KW Iron storage; Iron; Metal-binding.
FT INIT MET 0 0 BY SIMILARITY.
FT DOMAIN 6 155 FERRITIN-LIKE DIIRON.
FT METAL 53 53 IRON (POTENTIAL).
FT METAL 56 56 IRON (POTENTIAL).
FT METAL 56 56 IRON (POTENTIAL).
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 16:54:27 ; Search time 4.66667 Seconds
(without alignments)
90.694 Million cell updates/sec

Title: US-10-014-658-2

Perfect score: 42

Sequence: 1 STEVEAAGR 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	33	78.6	456	1	TRME LEPIN
2	33	78.6	1376	1	VCAP_HSVB
3	32	76.2	459	1	ZPR1_HUMAN
4	32	76.2	459	1	ZPR1_MOUSE
5	31	73.8	174	1	FRIL_BOVIN
6	31	73.8	174	1	FRIL_HORSE
7	31	73.8	182	1	FRIL_RAT
8	31	73.8	182	1	FRIL_MOUSE
9	31	73.8	182	1	FRIL_MOUSE
10	31	73.8	380	1	VGL1_HSVB
11	31	73.8	2499	1	MPRI_BOVIN
12	30	71.4	113	1	RA2B_MAIZE
13	30	71.4	432	1	CBAA_COMTE
14	30	71.4	453	1	SOXA_RHOS
15	30	71.4	501	1	GLPK_DEIRA
16	30	71.4	548	1	PNK1_MOUSE
17	30	71.4	642	1	PRIM_MYCLE
18	30	71.4	753	1	IDEP_ECOLI
19	30	71.4	1059	1	CAPU_DROME
20	30	71.4	1520	1	ABL_DROME
21	30	71.4	2366	1	TOXB_CLODI
22	30	71.4	2483	1	MPRI_MOUSE
23	29.5	70.2	499	1	DHA2_HUMAN
24	29.5	70.2	499	1	DHA2_MOUSE
25	29.5	70.2	499	1	DHA2_RAT
26	29.5	70.2	499	1	DHAS_CHICK
27	29	69.0	245	1	TRUA_ZYMO
28	29	69.0	251	1	TRPC_HALVO
29	29	69.0	309	1	ARCL_MYCPN
30	29	69.0	312	1	ARC2_ENTFA
31	29	69.0	441	1	DHOM_MYCLE
32	29	69.0	441	1	DHOM_MYCTU
33	29	69.0	445	1	DHOM_CORGL

34	29	69.0	456	1	MURD_ENTFA
35	29	69.0	510	1	GPNI_VIBCH
36	29	69.0	556	1	THSB_HALNI
37	29	69.0	567	1	ILVD_STREN
38	29	69.0	611	1	SCIA_ARATH
39	29	69.0	785	1	SCIL_SCHPO
40	29	69.0	788	1	TRSL_HCMVA
41	29	69.0	846	1	IRSL_HCMVA
42	29	69.0	940	1	GLND_CAUCR
43	29	69.0	1420	1	SRE9_YEAST
44	29	69.0	4303	1	PKD1_HUMAN
45	28.5	67.9	161	1	PHCA_SYNFW

ALIGNMENTS

RESULT 1

ID	TRME LEPIN	STANDARD;	PRT;	456 AA.
AC	P97043;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Probable tRNA modification GTPase trmE.			
GN	TRME OR LA0180.			
OS	Leptospiira interrogans.			
OC	Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.			
OX	NCBI_TaxID=173;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Ictero No.1 / Serogroup Icterohaemorrhagiae;			
RX	MEDLINE=98332717; PubMed=566070;			
RA	Takanashi Y., Akase K., Hirano H., Fukunaga M.;			
RT	Physical and genetic maps of the Leptospiira interrogans serovar			
RT	icterohaemorrhagiae strain Ictero No. 1 chromosome and sequencing of a			
RT	19-kb region of the genome containing the 5S rRNA gene."			
RL	Gene 215:37-45(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar lai;			
RX	MEDLINE=22598143; PubMed=12712204;			
RA	Ren S.-X., Fu G., Jiang X.-G., Zeng R., Miao Y.-G., Xu H.,			
RA	Zhang Y.-X., Xiong H., Lu G., Lu L.-F., Jiang H.-Q., Jia J., Tu Y.-F.,			
RA	Jiang J.-X., Gu W.-Y., Zhang Y.-Q., Cai Z., Sheng H.-H., Yin H.-F.,			
RA	Zhang Y., Zhu G.-F., Wan M., Huang H.-L., Qian Z., Wang S.-Y., Ma W.,			
RA	Yao Z.-J., Shen Y., Qiang B.-Q., Xia Q.-C., Guo X.-K., Danchin A.,			
RA	Saint Girons I., Somerville R.L., Wen Y.-N., Shi M.-H., Chen Z.,			
RA	Xu J.-G., Zhao G.-P.;			
RT	"Unique physiological and pathogenic features of Leptospiira			
RT	interrogans revealed by whole-genome sequencing."			
RL	Nature 422:898-993(2003).			
CC	-1- FUNCTION: EXHIBITS A VERY HIGH INTRINSIC GTPASE HYDROLYSIS RATE.			
CC	INVOLVED IN THE BIOSYNTHESIS OF THE HYPERMODIFIED NUCLEOSIDE 5-			
CC	METHYLAMINOMETHYL-2-THIOURINE, WHICH IS FOUND IN THE WOBBLE			
CC	POSITION OF SOME TRNAS (BY SIMILARITY).			
CC	-1- SIMILARITY: Belongs to the era/trmE family of GTP-binding			
CC	proteins. TrmE subfamily.			

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

EMBL; AB001721; BAAL9450.1; ALT_INIT.
EMBL; AB010203; BAA24373.1; ALT_INIT.
EMBL; AB011208; AAN47379.1; -.
PIR; T00126; T00126.
HMAP; MF_00379; -; 1.
InterPro; IPR005289; Gtp-binding_dom.

RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Uterback T., Tran K., Wolf A., Vanathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of *Caulobacter crescentus*.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005968; AAK25017.1; -.
DR TIGR; CC3055; -.
DR InterPro; IPR002195; Dihydroorotase.
DR PROSITE; PS00482; DIHYDROOROTASE_1; 1.
RW Hypothetical protein; Complete proteome.
SQ SEQUENCE 738 AA; 73548 MW; F99CDE793D68D63A CRC64;

Query Match 78.6%; Score 33; DB 16; Length 738;
Best Local Similarity 100.0%; Pred. No. 3.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 EYEAAAGR 9
Db 465 EYEAAAGR 471
|||||

RESULT 12
Q9RKE3 PRELIMINARY; PRT; 225 AA.
ID Q9RKE3
AC Q9RKE3
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Putative aldolase.
GN SC03495 OR SCE65.31C.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Murphy L., Harris D.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapate D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.B., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL939116; CAB61819.1; -.
DR InterPro; IPR000887; Aldolase_KDG_KHG.
DR Pfam; PF01081; Aldolase; 1.

DR TIGRPFAMs; TIGR01182; eda; 1.
KW Complete proteome.
SQ SEQUENCE 225 AA; 22578 MW; B0DFOE4F96343162 CRC64;

Query Match 76.2%; Score 32; DB 16; Length 225;
Best Local Similarity 75.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TEVEAAGR 9
Db 128 TEIHAAGR 135
|||||

RESULT 13
Q8C8U3 PRELIMINARY; PRT; 426 AA.
ID Q8C8U3
AC Q8C8U3
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Zinc finger protein 259.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Retina;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK044472; BAC31939.1; -.
SQ SEQUENCE 426 AA; 46786 MW; B0BC79D1E22D29B9 CRC64;

Query Match 76.2%; Score 32; DB 11; Length 426;
Best Local Similarity 55.6%; Pred. No. 3.4e+02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 STEVEAAGR 9
Db 87 NTEIQSAGR 95
|||||

RESULT 14
Q9HT33 PRELIMINARY; PRT; 445 AA.
ID Q9HT33
AC Q9HT33
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Probable dihydroorotase.
GN PA5541.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of *Pseudomonas aeruginosa* PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL; AE004966; AAG08926.1; -.

```

Best Local Similarity 100.0%; Pred. No. 1.1e+02; Mismatches 0; Indels 0; Gaps 0;
Matches 7; Conservative 0;

QY 3 EVEAAGR 9
Db 95 EVEAAGR 101

RESULT 6
Q8ST63
ID Q8ST63 PRELIMINARY; PRT; 274 AA.
AC Q8ST63
RX MEDLINE=21576510; PubMed=11719806;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein XAC0555.
GN XAC0555.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
RP STRAIN=306 / ATCC 13902 / XV 101;
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=2202145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furian L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Cannavan F., Cardoso J., Chambergo F., Ciapina L.P.,
RA Ciccarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locati E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463(2002).
DR EMBL; AB011683; AM35444.1; --
DR InterPro; IPR000104; Antifreeze_1.
DR PRINTS; PRO0308; ANTIFREEZE1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 274 AA; 28087 MW; 64B012849B5ED56D CRC64;

Query Match 78.6%; Score 33; DB 16; Length 274;
Best Local Similarity 87.5%; Pred. No. 1.4e+02;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 STEVEAAG 8
Db 178 STKVEAAG 185

RESULT 7
Q8ST63
ID Q8ST63 PRELIMINARY; PRT; 280 AA.
AC Q8ST63
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein ECU07_1810 (Hypothetical protein
DE ECU10_0080).
DN ECU07_1810 OR ECU10_0080.
OS Encephalitozoon cuniculi.
OC Eukaryota; Fungi; Microsporidia; Unikaryonidae; Encephalitozoon.
OX NCBI_TaxID=6035;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=GB-M1;
RA Genoscope;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=GB-M1;
RX MEDLINE=21576510; PubMed=11719806;
RA Katinka M.D., Duprat S., Cornillot E., Metenier G., Thomarar F.,
RA Prensier G., Barbe V., Peyretailade E., Brottier P., Wincker P.,
RA Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M.,
RA Weissenbach J., Vivares C.P.;
RT "Genome sequence and gene compaction of the eukaryote parasite
RT Encephalitozoon cuniculi."
RL Nature 414:450-453(2001).
DR EMBL; AL590447; CAD25712.1; --
DR EMBL; AL590449; CAD25728.1; --
KW Hypothetical protein.
SQ SEQUENCE 280 AA; 29871 MW; D87A380A895D1C35 CRC64;

Query Match 78.6%; Score 33; DB 5; Length 280;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 EVEAAGR 9
Db 242 EVEAAGR 248

RESULT 8
Q8ST22
ID Q8ST22 PRELIMINARY; PRT; 280 AA.
AC Q8ST22
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein ECU11_2050.
GN ECU11_2050.
OS Encephalitozoon cuniculi.
OC Eukaryota; Fungi; Microsporidia; Unikaryonidae; Encephalitozoon.
OX NCBI_TaxID=6035;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GB-M1;
RX MEDLINE=21576510; PubMed=11719806;
RA Katinka M.D., Duprat S., Cornillot E., Metenier G., Thomarar F.,
RA Prensier G., Barbe V., Peyretailade E., Brottier P., Wincker P.,
RA Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M.,
RA Weissenbach J., Vivares C.P.;
RT "Genome sequence and gene compaction of the eukaryote parasite
RT Encephalitozoon cuniculi."
RL Nature 414:450-453(2001).
DR EMBL; AL590450; CAD26115.1; --
KW Hypothetical protein.
SQ SEQUENCE 280 AA; 29837 MW; D0D8BA0A895D1C3B CRC64;

Query Match 78.6%; Score 33; DB 5; Length 280;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 EVEAAGR 9
Db 242 EVEAAGR 248

RESULT 9
Q31231
ID Q31231 PRELIMINARY; PRT; 349 AA.
AC Q31231;

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 18:03:30 ; Search time 20.3333 Seconds
(without alignments)
114.220 Million cell updates/sec

Title: US-10-014-658-2

Perfect score: 42

Sequence: 1 STEVEAAGR 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL.23.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	35	83.3	362	16 Q88NU2	Q88nu2 rhizobium 1
2	35	83.3	1020	4 Q8WU84	Q8wu84 homo sapien
3	34	81.0	479	16 Q8ZNU1	Q8znu1 salmonella
4	33	78.6	110	16 Q8YM01	Q8ym01 arabesna sp
5	33	78.6	223	16 Q988D5	Q988d5 rhizobium 1
6	33	78.6	274	16 Q8PFX8	Q8ppx8 xanthomonas
7	33	78.6	280	5 Q8ST63	Q8st63 encephalito
8	33	78.6	280	5 Q8ST22	Q8st22 encephalito
9	33	78.6	349	2 Q31231	Q31231 arthrobacte
10	33	78.6	349	2 Q8GAG7	Q8gag7 arthrobacte
11	33	78.6	738	16 Q9A3Z5	Q9a3z5 caulobacter
12	32	76.2	225	16 Q9RKE3	Q9rke3 streptomyce
13	32	76.2	426	11 Q8C8U3	Q8c8u3 mus musculu
14	32	76.2	445	16 Q9HT33	Q9ht33 pseudomonas
15	32	76.2	459	11 Q9JJA1	Q9jja1 mus musculu
16	32	76.2	913	10 Q9M8Y2	Q9m8y2 arabidopsis

Q8h3Y0 oryza sativ
Q8t5J8 anopheles g
Q8cJn6 streptomyce
Q8cwa4 mus musculu
Q86877 streptococc
Q84541 streptococc
Q8jkm8 cavia porce
Q8jpk6 cavia porce
Q8cpk4 mus musculu
Q8dy19 bovine herp
Q8rbs5 thermocane
Q84lp3 oryza sativ
Q89x15 streptococc
Q8nyz8 streptococc
Q8s127 agrobacteri
Q89z18 streptococc
Q8aab9 caulobacter
Q84829 streptococc
Q8g508 bifidobacte
Q8kgt4 agrobacteri
P1952 mycobacteri
Q8z5t6 salmonella
Q8s1k1 caenorhabdi
Q8s180 arabidopsi
Q8n0m8 arabidopsi
Q8sq25 ovine aries
Q8zev8 versinia pe
Q85mi8 cynocephalu
Q85mj1 lemur catia

ALIGNMENTS

RESULT 1

Q98NU2 PRELIMINARY; PRT; 362 AA.
ID AC Q98NU2
AC Q98NU2
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE Replication protein A.
GN MUR9740.
OS Rhizobium loti (Mesorhizobium loti).
OG Plasmid pMLb.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Watanabe A., Idesawa Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Kaneko T., Nakamura Y., Sato S., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL: AP003017; BAB54913.1; --
KW Plasmid; Complete Proteome.
SQ SEQUENCE 362 AA; 49812 MW; 71B60FA225769B55 CRC64;

Query Match 83.3%; Score 35; DB 16; Length 362;
Best Local Similarity 87.5%; Pred. No. 71;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 2 TEVEAAGR 9

Db 271 TEVEASGR 278

PT treating acne vulgaris -

XX Example 1; SEQ ID No 13803; 1069pp; English.

XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic polypeptides. The proteins and their associated DNA sequences are used in the treatment, prevention and diagnosis of medical conditions caused by P. acnes. The disorders include SAPHO syndrome (synovitis, acne, pustulosis, hypertyosis and osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved in infections of bone, joints and the central nervous system, however it is particularly involved in the inflammatory lesions associated with acne vulgaris. A method for detecting the presence or absence of P. acnes in a patient comprises contacting a sample with a binding agent that binds to the proteins of the invention and determining the amount of bound protein in the sample. The polypeptides may be used as antigens in the production of antibodies specific for P. acnes proteins. These antibodies can be used to downregulate expression and activity of P. acnes polypeptides and therefore treat P. acnes infections. The antibodies may also be used as diagnostic agents for determining P. acnes presence, for example, by enzyme linked immunosorbent assay (ELISA).

XX Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX

SQ Sequence 66 AA;

Query Match 70.2%; Score 33; DB 22; Length 66;

Best Local Similarity 77.8%; Pred. No. 49;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 STEGPFSGR 9

DB 6 SLERPFSGR 14

RESULT 15

AAW26430

ID AAW26430 standard; Protein; 92 AA.

XX

AC AAW26430;

XX

DT 25-MAR-2003 (updated)

DT 16-DEC-1997 (first entry)

XX

DE Swinepox virus HindIII C encoded protein C20L.

XX

XX SPV; vaccine; vector; pseudorabies virus; C20L.

XX

OS Swinepox virus.

XX

PN US5651972-A.

XX

FD 29-JUL-1997.

XX

XX 14-SEP-1994; 94US-0307499.

XX

PR 01-JUL-1992; 92US-0908241.

PR 21-APR-1989; 89US-0342212.

PR 23-JUN-1992; 92US-0308630.

PR 14-SEP-1994; 94US-0307499.

XX

XX (UYFL) UNIV FLORIDA RES FOUND INC.

XX

XX Gibbs EPJ, Moyer RW, Vinuela E;

XX

DR WPI; 1997-392897/36.

DR N-PSDB; AAT84564.

XX

XX Recombinant swinepox virus vector - used particularly for vaccines against infectious agents, including pseudorabies

XX

XX Example 2; Column 101-104; 70pp; English.

XX

XX This sequence comprises a polypeptide, designated C20L, encoded by the antisense strand (open reading frame 1) of the HindIII C fragment (see AAT84564) of swinepox virus (SPV). C20L shows 72% identity with vaccinia putative protein kinase (432 amino acid (aa) overlap) and 20% identity with yeast clathrin heavy chain protein (54 aa overlap). A claimed recombinant vector comprises a heterologous nucleotide sequence inserted into, or replacing, all or a portion of a non-essential SPV gene or nucleic acid sequence of the HindIII C fragment. The vector can be used for the expression of heterologous proteins, both in vivo as a vaccine, and in vitro for production of the selected protein. The heterologous protein is preferably pseudorabies virus gp50 or gp63 for use in swine vaccines. As SPV is host-restricted to swine, the use of modified recombinant SPV as a live vaccine vector eliminates the risk of spreading infection with the virus to other animal populations.

XX (Updated on 25-MAR-2003 to correct PF field.)

XX

SQ Sequence 92 AA;

Query Match 70.2%; Score 33; DB 18; Length 92;

Best Local Similarity 100.0%; Pred. No. 89;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 EGPFSG 8

DB 84 EGPFSG 89

Search completed: December 11, 2003, 18:24:39

Job time : 27 secs

PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX PA
 XX Rosen CA, Barash SC, Ruben SM;
 PI N-PSDB; AAS41063.
 DR WPI; 2001-465566/50.
 DR N-PSDB; AAS41063.
 XX Novel polypeptides and polynucleotides useful for diagnosing,
 PT preventing, treating neural, immune system, muscular, reproductive,
 PT Pulmonary, cardiovascular, renal, proliferative disorders and cancerous
 PT diseases -
 XX
 XX Claim 11; SEQ ID No 1189; 1180pp; English.
 XX The present invention relates to the isolation of novel human enzyme
 CC polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences
 CC encoding them. The enzyme polypeptides of the invention may comprise the
 CC functional classes of oxidoreductases, transferases, hydrolases, lyases,
 CC isomerases or ligases. The sequences of the invention are useful in the
 CC diagnosis, treatment, prevention and/or prognosis of a wide range of
 CC disorders including hyperproliferative disorders (e.g. cancer),
 CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders
 CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
 CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders
 CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
 CC blood-related disorders (e.g. haemophilia), reproductive disorders
 CC (e.g. infertility) and infectious disorders (e.g. influenza). The
 CC polynucleotides of the invention can also be used in gene therapy.
 CC AAU22915-AAU23814 represent the novel human enzyme polypeptides of the
 CC invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX SQ Sequence 336 AA;
 Query Match 72.3%; Score 34; DB 22; Length 336;
 Best Local Similarity 75.0%; Pred. No. 1.7e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 Qy 2 TEGFFSGR 9
 |||||:
 188 TEGFIAGR 195
 Db
 RESULT 11
 AAW79282
 ID AAW79282 standard; Protein; 561 AA.
 XX
 XX AC AAW79282;
 XX
 XX 15-FEB-1999 (first entry)
 DT
 XX Tomato ringspot virus peach isolate coat protein.
 DE TRSV; coat protein; transgenic plant; disease resistance.
 XX
 XX OS Tomato ringspot virus.
 XX WO9844803-A1.
 PN 15-OCT-1998.
 FD 03-APR-1998; 98WO-US06726.
 XX
 XX PF 04-APR-1997; 97US-0042658.
 XX

PA (CORR) CORNELL RES FOUND INC.
 XX Fuchs M, Gonsalves D, Yepes L;
 XX WPI; 1998-594471/50.
 DR N-PSDB; AAV65390, AAV65391.
 XX Isolated coat protein from tomato ringspot virus and related DNA,
 PT vectors, transformed cells - antibodies and transgenic plants with
 PT inheritable resistance to the virus
 XX
 XX Claim 2; Page 9-11; 65pp; English.
 XX This is the amino acid sequence of tomato ringspot virus (TRSV)
 CC coat protein (see AAW79282), as deduced from a coding sequence (see
 CC AAV65391) in the 3' terminal region of the RNA 2 of a peach isolate
 CC of TRSV. The invention provides an isolated coat protein or
 CC polypeptide from a peach isolate of TRSV, and a DNA sequence
 CC encoding it, as well as a heterologous expression system, host
 CC cells, transgenic plants containing coat protein DNA or its
 CC antisense form (also their seeds and propagules), antibodies and
 CC probes that recognise the coat protein. The transgenic plants
 CC (specifically apple, peach, grape, raspberry, cherry, plum and
 CC strawberry) are almost immune to TRSV, even though they do not
 CC produce viral coat protein. Since coat protein is not produced,
 CC there is no danger that heterocapsidation or recombination
 CC between transgenic and viral RNAs can occur. TRSV can be diagnosed
 CC using the antibodies or probes.
 XX SQ Sequence 561 AA;
 Query Match 72.3%; Score 34; DB 19; Length 561;
 Best Local Similarity 66.7%; Pred. No. 2.9e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 STEGFFSGR 9
 |||||:
 396 SSTGFTGR 404
 Db
 RESULT 12
 AEG18170
 ID AEG18170 standard; Protein; 796 AA.
 XX
 XX AC AEG18170;
 XX
 XX 18-FEB-2002 (first entry)
 DT
 XX Novel human diagnostic protein #18161.
 DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 KW Homo sapiens.
 OS WO200175067-A2.
 XX
 XX PD 11-OCT-2001.
 XX
 XX 30-MAR-2001; 2001WO-US08631.
 PF
 XX 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX (HYSE-) HYSEQ INC.
 XX Drmanac RT, Liu C, Tang YT;
 PI WPI; 2001-639362/73.
 XX N-PSDB; AAS82357.
 DR New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT

RESULT 8
ABG29070
ID ABG29070 standard; Protein; 286 AA.
XX
AC
XX
AC
XX
ABG29070;
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #29061.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
PA
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS93257.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX
PS Claim 20; SEQ ID No 59429; 103pp; English.

PS Claim 20; SEQ ID No 59429; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG0010-ABG3037 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 286 AA;

Query Match 72.3%; Score 34; DB 22; Length 286;
Best Local Similarity 75.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 TEGFFPSGR 9
DB 153 TEGFIAGR 160

RESULT 9
AAG92009

ID AAG92009 standard; Protein; 321 AA.
XX
AC
XX
AC
XX
AAG92009;
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum protein fragment SEQ ID NO: 5763.
XX
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis.
XX
OS Corynebacterium glutamicum.
XX
PN EF1108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
XX (KYOW) KYOWA HAKKO KOGYO KK.
XX
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A,
XX
XX WPI; 2001-376931/40.
XX N-PSDB; AAH67228.

XX Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT
XX
XX

PS Claim 17; SEQ ID NO: 5763; 246pp + Sequence Listing; English.

XX The present invention provides a number of nucleotide and protein
XX sequences from the Coryneform bacterium Corynebacterium glutamicum. These
XX are useful for identifying the mutation point of a gene derived from a
XX mutant of coryneform bacterium, measuring expression amount and
XX analysing the expression profile or expression pattern of a gene derived
XX from Coryneform bacterium, and identifying a homologue of a gene derived
XX from Coryneform bacterium. Coryneform bacteria are useful for producing
XX amino acids, nucleic acids, vitamins, saccharides and organic acids,
XX particularly L-lysine. The present sequence is a protein described
XX in the exemplification of the invention.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from the
XX European Patent Office.

XX Sequence 321 AA;

Query Match 72.3%; Score 34; DB 22; Length 321;
Best Local Similarity 85.7%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 EGFPSGR 9
DB 275 EGFPSGR 281

RESULT 10
AAU23193
ID AAU23193 standard; Protein; 336 AA.

XX
AC AAU23193;
XX
XX
DT 17-DEC-2001 (first entry)
XX
DE Novel human enzyme polypeptide #279.

CC AAC78457 and AA844240 represent sequences used in the exemplification of
CC the present invention.

SQ Sequence 127 AA;

Query Match 74.5%; Score 35; DB 21; Length 127;
Best Local Similarity 85.7%; Pred. No. 39;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 EGFPSGR 9
|||:
DB 2 EGFYSGR 8

RESULT 6

AAU23731
ID AAU23731 standard; Protein; 117 AA.

XX AC

AAU23731;

XX DT 18-DEC-2001 (first entry)

XX DE

XX Novel human enzyme polypeptide #817.

XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;
KW nephrotropic; anticoagulant.

XX OS Homo sapiens.

XX PN WO20015301-A2.

XX PD 02-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US01239.

XX 31-JAN-2000; 2000US-0179065.

XX 04-FEB-2000; 2000US-0180628.

XX 24-FEB-2000; 2000US-0184664.

XX 02-MAR-2000; 2000US-0186350.

XX 16-MAR-2000; 2000US-0189874.

XX 17-MAR-2000; 2000US-0190076.

XX 18-APR-2000; 2000US-0198123.

XX 19-MAY-2000; 2000US-0205515.

XX 07-JUN-2000; 2000US-0209467.

XX 28-JUN-2000; 2000US-0214886.

XX 30-JUN-2000; 2000US-0215135.

XX 07-JUL-2000; 2000US-0216647.

XX 07-JUL-2000; 2000US-0216880.

XX 11-JUL-2000; 2000US-0217487.

XX 11-JUL-2000; 2000US-0217496.

XX 14-JUL-2000; 2000US-0218290.

XX 26-JUL-2000; 2000US-0220963.

XX 26-JUL-2000; 2000US-0220964.

XX 14-AUG-2000; 2000US-0224518.

XX 14-AUG-2000; 2000US-0224519.

XX 14-AUG-2000; 2000US-0225213.

XX 14-AUG-2000; 2000US-0225214.

XX 14-AUG-2000; 2000US-0225266.

XX 14-AUG-2000; 2000US-0225267.

XX 14-AUG-2000; 2000US-0225268.

XX 14-AUG-2000; 2000US-0225270.

XX 14-AUG-2000; 2000US-0225447.

XX 14-AUG-2000; 2000US-0225757.

XX 14-AUG-2000; 2000US-0225758.

XX 14-AUG-2000; 2000US-0225759.

XX 18-AUG-2000; 2000US-0226279.

XX 22-AUG-2000; 2000US-0226681.

XX 22-AUG-2000; 2000US-0226688.

XX 22-AUG-2000; 2000US-0227182.

PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0232403.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 25-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235936.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241321.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.

PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0228868.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0232403.
 PR 14-SEP-2000; 2000US-0232404.
 PR 14-SEP-2000; 2000US-0233064.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234977.
 PR 25-SEP-2000; 2000US-0234998.
 PR 25-SEP-2000; 2000US-0234999.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239335.
 PR 13-OCT-2000; 2000US-0239337.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.

PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249254.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251858.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-465460/50.

N-PSDB; RAS27383.

Novel polypeptides useful for diagnosing, treating, preventing and/or
 prognosing disorders related to the proteins, including cancers, immune
 disorders and neuronal disorders -

Claim 1; SEQ ID No 1031; 880pp; English.

The invention relates to novel isolated polypeptides (I), and
 polynucleotides (II). (I), (II) and the antibody to (I) are useful for
 diagnosing, preventing and treating diseases including immune system
 disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
 disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ
 transplant rejections and graft versus host disease, infectious diseases
 (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 17:22:25 ; Search time 25 Seconds
(without alignments)
57,142 Million cell updates/sec

Title: US-10-014-658-4

Perfect score: 47

Sequence: 1 STEGFFSGR 9

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_13Jun03.*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	47	100.0	9	21	AAU44468 Human antithrombin
2	38	80.9	679	20	AAU35714 Amino acid sequenc
3	36	76.6	194	22	AAU17466 Novel signal trans
4	36	76.6	200	21	AAU58673 Porcine reproducti
5	35	74.5	127	21	AAU43960 Human cancer assoc
6	34	72.3	117	22	AAU23731 Novel human enzyme
7	34	72.3	286	22	ABG28991 Novel human diagno
8	34	72.3	286	22	ABG29070 Novel human diagno
9	34	72.3	321	22	ABG29009 C glutamicum prote

10	34	72.3	336	22	AAU23193 Novel human enzyme
11	34	72.3	561	19	AAW73282 Tomato ringspot vi
12	34	72.3	796	22	ABG18170 Novel human diagno
13	34	72.3	896	22	ABG28984 Novel human diagno
14	33	70.2	66	22	AAU52608 Propionibacterium
15	33	70.2	92	18	AAW26430 Swinepox virus Hin
16	33	70.2	92	22	AAU68251 Protein encoded by
17	33	70.2	112	21	AAU40761 Human ORF525
18	33	70.2	115	21	AAU40736 Human ORF500
19	33	70.2	115	23	ABP33559 Human ORF4332 Prot
20	33	70.2	117	21	AAU41979 Human ORF1743
21	33	70.2	137	22	AAU67051 Human dihydroorota
22	33	70.2	143	22	AAU24117 Human EST encoded
23	33	70.2	191	23	AAE15938 Human 5729744 PNAS
24	33	70.2	193	23	AAE15939 Mouse 4191392 PNAS
25	33	70.2	194	20	AAU37112 Protein involved i
26	33	70.2	195	22	ABG23225 Novel human diagno
27	33	70.2	225	21	AAU87235 Human signal pepti
28	33	70.2	295	21	AAU51904 Murine 3GnT protei
29	33	70.2	299	21	AAU51903 Human 3GnT protei
30	33	70.2	305	21	AAU55049 Arabidopsis thalia
31	33	70.2	311	20	AAU34913 Chlamydia pneumoni
32	33	70.2	311	23	ABP60854 Chlamydia pneumoni
33	33	70.2	312	23	ABP60853 Chlamydia muridaru
34	33	70.2	312	23	ABP60855 Chlamydia trachoma
35	33	70.2	317	23	ABU47589 Listeria monocytog
36	33	70.2	325	21	AAU51902 Murine 3GnT protei
37	33	70.2	329	21	AAU51901 Human 3GnT protei
38	33	70.2	329	24	ABP60071 B3GALT polypeptide
39	33	70.2	329	22	ABG04074 Novel human diagno
40	33	70.2	361	21	ABG30297 Arabidopsis thalia
41	33	70.2	384	21	ABG30296 Arabidopsis thalia
42	33	70.2	390	21	ABG30295 Arabidopsis thalia
43	33	70.2	390	23	ABU91529 Herbicidally activ
44	33	70.2	409	22	AAE00710 Rhodospseudomonas s
45	33	70.2	489	24	AAE29805 Streptococcus sp.

ALIGNMENTS

RESULT 1

AAU44468

ID AAU44468 standard; peptide; 9 AA.

XX AAU44468;

AC AAU44468;

XX 27-MAR-2000 (first entry)

DT 27-MAR-2000 (first entry)

XX Human antithrombin III variant Bb (residues 385-393).

XX Human; antithrombin III; ATIII variant Bb; elastase-resistant;

XX IGG activated neutrophil resistant; anti-thrombin activity; heparin;

XX anti-factor Xa activity; blood clotting disorder; sepsis; trauma; stroke;

XX thrombin activation-related pathological symptom; restenosis; thrombosis;

XX acute respiratory distress syndrome; thromboembolism; reocclusion.

XX Homo sapiens.

OS Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

XX Synthetic.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run On: December 11, 2003, 18:21:27 ; Search time 49.3333 Seconds
(without alignments)
33.929 Million cell updates/sec

Title: US-10-014-658-1
Perfect score: 41
Sequence: 1 STALEAIG 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 684280 seqs, 185983659 residues

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/prodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/prodata/1/pubpaa/US05_NEW_PUB.pep.*
- 4: /cgn2_6/prodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/prodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pep.*
- 7: /cgn2_6/prodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/prodata/1/pubpaa/US09_PUBCOMB.pep.*
- 9: /cgn2_6/prodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/prodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/prodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/prodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/prodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/prodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/prodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/prodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/prodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/prodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	32	78.0	452	15	US-10-156-761-15025
2	32	78.0	575	9	US-09-970-711-30
3	32	78.0	974	15	US-10-128-714-8018
4	31	75.6	247	11	US-09-880-505-135
5	31	75.6	247	14	US-10-051-643-135
6	31	75.6	259	10	US-09-738-626-3543
7	31	75.6	299	11	US-09-880-505-124
8	31	75.6	299	12	US-09-809-391-432
9	31	75.6	299	12	US-09-882-171-432
10	31	75.6	299	14	US-10-051-643-124
11	31	75.6	476	15	US-10-156-761-11391
12	31	75.6	670	11	US-09-880-505-178
13	31	75.6	670	12	US-10-205-979-47
14	31	75.6	670	14	US-10-051-643-178
15	31	75.6	722	11	US-09-880-505-174

Sequence 174, App	722	14	US-10-051-643-174
Sequence 1082, Ap	75.6	31	US-10-238-075-1082
Sequence 76, Appl	1116	12	US-09-917-340-76
Sequence 4381, Ap	95	9	US-09-738-626-4381
Sequence 19, Appl	200	10	US-10-269-557-19
Sequence 9816, Ap	502	15	US-10-156-761-9816
Sequence 20, Appl	534	15	US-10-269-557-20
Sequence 21, Appl	550	15	US-10-269-557-21
Sequence 3097, Ap	570	15	US-10-128-714-3097
Sequence 8097, Ap	586	15	US-10-128-714-8097
Sequence 53, Appl	587	15	US-10-267-311-53
Sequence 23, Appl	647	15	US-10-060-230-23
Sequence 24, Appl	736	14	US-10-060-230-24
Sequence 5568, Ap	736	14	US-09-815-243-5568
Sequence 12162, A	1491	9	US-09-815-243-12162
Sequence 15, Appl	1502	9	US-10-227-639-15
Sequence 4839, Ap	37	15	US-09-738-626-4839
Sequence 12388, A	271	10	US-10-156-761-12388
Sequence 232, App	289	15	US-09-934-455-232
Sequence 10908, A	381	11	US-10-156-761-10908
Sequence 2, Appl	417	15	US-09-950-788-2
Sequence 4, Appl	568	10	US-09-950-788-4
Sequence 7, Appl	568	10	US-09-950-788-7
Sequence 5622, Ap	568	10	US-09-738-626-5622
Sequence 2, Appl	568	12	US-10-460-294-2
Sequence 4, Appl	568	12	US-10-460-294-4
Sequence 7, Appl	568	12	US-10-460-294-7
Sequence 20, Appl	735	12	US-10-381-779-20
Sequence 5468, Ap	735	12	US-09-738-626-5468
Sequence 4, Appl	831	10	US-10-153-219-4
Sequence 4, Appl	1107	15	US-10-153-219-4

ALIGNMENTS

RESULT 1
US-10-156-761-15025
; Sequence 15025, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 15025
; LENGTH: 452
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-15025

Query Match 78.0%; Score 32; DB 15; Length 452;
Best Local Similarity 75.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 STALEAIG 8
Db 357 TTALAEAVG 364

RESULT 2
US-09-970-711-30

```
RESULT 6
US-09-738-626-3543
; Sequence 3543, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 3543
; LENGTH: 259
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-3543

Query Match 75.6%; Score 31; DB 10; Length 259;
Best Local Similarity 65.7%; Pred. No. 1.2e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 STALEAIGR 9
Db 41 STVLKALGR 49

RESULT 7
US-09-880-505-124
; Sequence 124, Application US/09880505
; Publication No. US20030007976A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Tan, Paul L.J.
; APPLICANT: Prestidge, Ross
; TITLE OF INVENTION: Methods and Compounds for the Treatment
; TITLE OF INVENTION: of Immunologically-Mediated Skin Disorders
; FILE REFERENCE: 11000.1007c2
; CURRENT APPLICATION NUMBER: US/09/880,505
; CURRENT FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 09/324,542
; PRIOR FILING DATE: 1999-06-02
; PRIOR APPLICATION NUMBER: US 08/997,080
; PRIOR FILING DATE: 1997-12-23
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 124
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Mycobacterium vaccae
US-09-880-505-124

Query Match 75.6%; Score 31; DB 11; Length 299;
Best Local Similarity 87.5%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 STALEAIG 8
Db 11 STALEAIG 8

RESULT 8
US-09-809-391-432
; Sequence 432, Application US/09809391
; Publication No. US20030049618A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 186 Human Secreted proteins
; FILE REFERENCE: P2002P2
; CURRENT APPLICATION NUMBER: US/09/809,391
; CURRENT FILING DATE: 2001-03-16
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 761
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 432
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: SITE
; LOCATION: (4)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (254)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (255)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (299)
; OTHER INFORMATION: Xaa equals stop translation
US-09-809-391-432

Query Match 75.6%; Score 31; DB 11; Length 299;
Best Local Similarity 75.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TALEAIGR 9
Db 219 TALELVGR 226

RESULT 9
US-09-882-171-432
; Sequence 432, Application US/09882171
; Publication No. US20030175858A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 186 Human Secreted proteins
; FILE REFERENCE: P2002P2
; CURRENT APPLICATION NUMBER: US/09/882,171
; CURRENT FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 09/809,391
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 09/149,476
; PRIOR FILING DATE: 1998-09-08
; PRIOR APPLICATION NUMBER: PCT/US98/04493
; PRIOR FILING DATE: 1998-03-06
; PRIOR APPLICATION NUMBER: 60/040,162
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/040,333
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/038,621
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/040,626
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/040,334
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/040,336
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/040,163
```


; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/901,127
; FILING DATE:
; APPLICATION NUMBER: US 07/908,241
; FILING DATE: 1-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/908,630
; FILING DATE: 29-JUN-1992
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/342,212
; FILING DATE: 21-APR-1992
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UF35.1.FWCCL
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 92 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-299-268-31

Query Match 70.2%; Score 33; DB 3; Length 92;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 EGPFSG 8
Db 84 EGPFSG 89

RESULT 6
US-09-252-991A-19971
; Sequence 19971, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 19971
; LENGTH: 226
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19971

Query Match 70.2%; Score 33; DB 4; Length 226;
Best Local Similarity 85.7%; Pred. No. 44;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TEGFFSG 8
Db 176 TEGFFHG 182

RESULT 7
US-09-198-452A-331
; Sequence 331, Application US/09198452A

; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Grifflais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragment thereof and uses thereof, in particular for the diagnosis, prevention of infection and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 331
; LENGTH: 311
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-09-198-452A-331

Query Match 70.2%; Score 33; DB 4; Length 311;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 EGPFSG 8
Db 33 EGPFSG 38

RESULT 8
US-09-587-066-2
; Sequence 2, Application US/09587066
; Patent No. 6210945
; GENERAL INFORMATION:
; APPLICANT: LUNNEN, KEITH D.
; APPLICANT: MORGAN, RICHARD D.
; APPLICANT: MEIXSELL, TIMOTHY
; APPLICANT: WILSON, GEORGEY G.
; TITLE OF INVENTION: METHOD FOR CLONING AND PRODUCING THE REAR RESTRICTION
; ENDONUCLEASE IN E. COLI AND PURIFICATION OF THE
; FILE REFERENCE: NEB-179
; CURRENT APPLICATION NUMBER: US/09/587,066
; CURRENT FILING DATE: 2000-06-02
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 409
; TYPE: PRT
; ORGANISM: rhodopseudomonas sphaeroides
US-09-587-066-2

Query Match 70.2%; Score 33; DB 3; Length 409;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 GPFSGR 9
Db 203 GPFSGR 208

RESULT 9
US-09-328-352-5503
; Sequence 5503, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Braton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5503
; LENGTH: 3892
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii

QY 3 EGFPSGR 9
| | | | |
Db 1 EGFPSAR 7

RESULT 12
PCT-US94-07644A-41
; Sequence 41, Application PC/TUS9407644A
; GENERAL INFORMATION:
; APPLICANT: COR Therapeutics, Inc.
; TITLE OF INVENTION: Platelet Glycoprotein V Gene and Uses
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Kourie and Crew
; STREET: 379 Lytton Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: US
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/07644A
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dow, Karen B.
; REGISTRATION NUMBER: 29,684
; REFERENCE/DOCKET NUMBER: 012418-003000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 326-2400
; TELEFAX: (415) 326-2422
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; HYPOTHEetical: NO
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..11
; OTHER INFORMATION: /note= "Amino acid sequence of the
; OTHER INFORMATION: human fibrinogen (fg) B-beta chain
; OTHER INFORMATION: thrombin
; OTHER INFORMATION: cleavage site."
; FEATURE:
; NAME/KEY: Region
; LOCATION: 7..8
; OTHER INFORMATION: /note= "Amino acid residues
; OTHER INFORMATION: identical to GPV."
; FEATURE:
; NAME/KEY: Region
; LOCATION: 11
; OTHER INFORMATION: /note= "Amino acid residue
; OTHER INFORMATION: identical to GPV."
PCT-US94-07644A-41

Query Match 68.1%; Score 32; DB 5; Length 11;
Best Local Similarity 85.7%; Pred. No. 2.9;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 EGFPSGR 9
| | | | |
Db 1 EGFPSAR 7

RESULT 13
US-08-212-433A-15
; Sequence 15, Application US/08212433A
; Patent No. 5538897
; GENERAL INFORMATION:

; APPLICANT: Yates, III, John R.
; APPLICANT: Eng, James K.
; TITLE OF INVENTION: USE OF MASS SPECTROMETRY FRAGMENTATION
; TITLE OF INVENTION: PATTERNS OF PEPTIDES TO IDENTIFY AMINO ACID SEQUENCES IN
; TITLE OF INVENTION: DATABASES
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Kourie and Crew
; STREET: One Market Plaza, Steuart St. Tower
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94105-1492
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/212,433A
; FILING DATE: 14-MAR-1994
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: Hughes, Richard L.
; REGISTRATION NUMBER: 31,264
; REFERENCE/DOCKET NUMBER: 16336-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-467-9600
; TELEFAX: 415-543-5043
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-212-433A-15

Query Match 68.1%; Score 32; DB 1; Length 14;
Best Local Similarity 85.7%; Pred. No. 3.7;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 EGFPSGR 9
| | | | |
Db 8 EGFPSAR 14

RESULT 14
US-08-900-660A-4
; Sequence 4, Application US/08900660A
; Patent No. 5876947
; GENERAL INFORMATION:
; APPLICANT: Kudryk, Bohdan J
; APPLICANT: Bini, Alessandra
; APPLICANT: Zhang, Jian-Zhong
; TITLE OF INVENTION: MONOSPECIFIC ANTIBODY REACTIVE
; TITLE OF INVENTION: WITH FIBRINOGEN AND
; TITLE OF INVENTION: FIBRINOPEPTIDE B
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hoffmann & Baron, LLP
; STREET: 350 Jericho Turnpike
; CITY: Jericho
; STATE: New York
; COUNTRY: USA
; ZIP: 11753
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
; COMPUTER: IBM compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WordPerfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/900,660A

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 18:21:27 / Search time 49.3333 Seconds
(without alignments)
33.929 Million cell updates/sec.

Title: US-10-014-658-4
Perfect score: 47
Sequence: 1 STEGFFSGR 9

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 684280 seqs, 185983659 residues

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36	76.6	194	10	US-09-764-868-1031
2	35	74.5	127	9	US-09-925-301-1405
3	34	72.3	321	10	US-09-738-626-5763
4	33	70.2	280	12	US-10-237-386-66
5	33	70.2	311	12	US-10-032-201B-203
6	33	70.2	312	12	US-10-032-201B-202
7	33	70.2	312	12	US-10-032-201B-204
8	32	68.1	11	12	US-10-212-499-41
9	32	68.1	14	11	US-09-828-326-20
10	32	68.1	14	15	US-10-142-935-11
11	32	68.1	15	15	US-10-142-935-8
12	32	68.1	15	15	US-10-137-354-54
13	32	68.1	20	15	US-10-142-935-5
14	32	68.1	20	15	US-10-142-935-6
15	32	68.1	25	15	US-10-142-935-4

16 68.1 47 15 US-10-178-213-279 Sequence 279, App
17 68.1 78 15 US-10-178-213-278 Sequence 278, App
18 68.1 283 15 US-10-156-761-8768 Sequence 8768, App
19 68.1 314 12 US-10-354-804-2 Sequence 2, Appll
20 68.1 355 9 US-09-826-212-14 Sequence 14, Appll
21 68.1 355 10 US-09-935-727-16 Sequence 16, Appll
22 68.1 355 15 US-10-186-643-14 Sequence 14, Appll
23 68.1 404 12 US-10-354-804-13 Sequence 13, Appll
24 68.1 488 11 US-09-919-039-33 Sequence 33, Appll
25 68.1 491 15 US-10-017-724-6 Sequence 2, Appll
26 68.1 503 15 US-10-103-140-2 Sequence 15028, A
27 68.1 631 15 US-10-156-761-15028 Sequence 196, App
28 68.1 890 15 US-10-171-311-196 Sequence 53, Appll
29 68.0 47 12 US-10-072-809A-53 Sequence 396, App
30 68.0 47 15 US-10-178-213-396 Sequence 411, App
31 68.0 47 15 US-10-178-213-411 Sequence 81, Appll
32 68.0 49 15 US-10-178-213-81 Sequence 395, App
33 68.0 76 15 US-10-178-213-395 Sequence 410, App
34 68.0 77 15 US-10-178-213-410 Sequence 22, Appll
35 68.0 78 12 US-10-072-809A-22 Sequence 80, Appll
36 68.0 84 15 US-10-178-213-80 Sequence 7, Appll
37 68.0 201 15 US-10-203-224-7 Sequence 13, Appll
38 68.0 349 9 US-09-826-212-13 Sequence 15, Appll
39 68.0 349 10 US-09-935-727-15 Sequence 13, Appll
40 68.0 349 15 US-10-186-643-13 Sequence 41, Appll
41 68.0 350 15 US-10-046-433-41 Sequence 164, App
42 68.0 519 11 US-09-934-455-164 Sequence 7587, App
43 68.0 687 12 US-10-032-585-7587 Sequence 7, Appll
44 68.0 696 12 US-10-288-556-7 Sequence 11297, A
45 63.8 116 9 US-09-815-242-11297

ALIGNMENTS

RESULT 1
US-09-764-868-1031
; Sequence 1031, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P7232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1031
; LENGTH: 194
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-868-1031

Query Match 76.6%; Score 36; DB 10; Length 194;
Best Local Similarity 77.8%; Pred. No. 24;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 STEGFFSGR 9
Db 60 ATEGFTSGR 68

RESULT 2
US-09-925-301-1405
; Sequence 1405, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10

```

; TITLE OF INVENTION: METHODS FOR THE PRODUCTION OF MULTIMERIC PROTEINS, AND RELATED
; FILE REFERENCE: 38814 351B
; CURRENT APPLICATION NUMBER: US/10/032,201B
; CURRENT FILING DATE: 2001-12-19
; NUMBER OF SEQ ID NOS: 313
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 202
; LENGTH: 312
; TYPE: PRT
; ORGANISM: Chlamydia muridarum
US-10-032-201B-202

Query Match      70.2%; Score 33; DB 12; Length 312;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3 EGPFSG 8
DB      33 EGPFSG 38

RESULT 7
US-10-032-201B-204
; Sequence 204, Application US/10032201B
; Publication No. US20030167524A1
; GENERAL INFORMATION:
; APPLICANT: Van Rooijen, Gijb
; APPLICANT: Deckers, Harm
; APPLICANT: Heifetz, Peter Bernard
; APPLICANT: Briggs, Steven
; APPLICANT: Dalmia, Bipin Kumar
; APPLICANT: Del Val, Greg
; APPLICANT: Zaplachinski, Steve
; APPLICANT: Moloney, Maurice
; TITLE OF INVENTION: METHODS FOR THE PRODUCTION OF MULTIMERIC PROTEINS, AND RELATED
; FILE REFERENCE: 38814 351B
; CURRENT APPLICATION NUMBER: US/10/032,201B
; CURRENT FILING DATE: 2001-12-19
; NUMBER OF SEQ ID NOS: 313
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 204
; LENGTH: 312
; TYPE: PRT
; ORGANISM: Chlamydia trachomatis
US-10-032-201B-204

Query Match      70.2%; Score 33; DB 12; Length 312;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3 EGPFSG 8
DB      33 EGPFSG 38

RESULT 8
US-10-212-499-41
; Sequence 41, Application US/10212499
; Publication No. US20030135036A1
; GENERAL INFORMATION:
; APPLICANT: Lanza, Francois
; APPLICANT: Phillips, David R.
; APPLICANT: Cazenave, Jean-Pierre
; TITLE OF INVENTION: Platelet Glycoprotein V Gene and Uses
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morgan Lewis & Bockius LLP
; STREET: 1800 M St., NW
; CITY: Washington
; STATE: DC
; COUNTRY: US

; TITLE OF INVENTION: METHODS FOR THE PRODUCTION OF MULTIMERIC PROTEINS, AND RELATED
; FILE REFERENCE: 38814 351B
; CURRENT APPLICATION NUMBER: US/10/032,201B
; CURRENT FILING DATE: 2001-12-19
; NUMBER OF SEQ ID NOS: 313
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 202
; LENGTH: 312
; TYPE: PRT
; ORGANISM: Chlamydia muridarum
US-10-032-201B-202

Query Match      70.2%; Score 33; DB 12; Length 312;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3 EGPFSG 8
DB      33 EGPFSG 38

RESULT 9
US-09-828-326-20
; Sequence 20, Application US/09828326
; Publication No. US2003005952A1
; GENERAL INFORMATION:
; APPLICANT: Chait, Brian T.
; APPLICANT: Beavis, Ronald
; APPLICANT: Wang, Rong
; APPLICANT: Kent, Stephen B.H.
; TITLE OF INVENTION: Method and Product for the Sequence
; DETERMINATION OF PEPTIDES USING A MASS SPECTROMETER
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Crew LLP
```

```

; TITLE OF INVENTION: METHODS FOR THE PRODUCTION OF MULTIMERIC PROTEINS, AND RELATED
; FILE REFERENCE: 38814 351B
; CURRENT APPLICATION NUMBER: US/10/032,201B
; CURRENT FILING DATE: 2001-12-19
; NUMBER OF SEQ ID NOS: 313
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 202
; LENGTH: 312
; TYPE: PRT
; ORGANISM: Chlamydia muridarum
US-10-032-201B-202

Query Match      70.2%; Score 33; DB 12; Length 312;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3 EGPFSG 8
DB      33 EGPFSG 38

RESULT 7
US-10-032-201B-204
; Sequence 204, Application US/10032201B
; Publication No. US20030167524A1
; GENERAL INFORMATION:
; APPLICANT: Van Rooijen, Gijb
; APPLICANT: Deckers, Harm
; APPLICANT: Heifetz, Peter Bernard
; APPLICANT: Briggs, Steven
; APPLICANT: Dalmia, Bipin Kumar
; APPLICANT: Del Val, Greg
; APPLICANT: Zaplachinski, Steve
; APPLICANT: Moloney, Maurice
; TITLE OF INVENTION: METHODS FOR THE PRODUCTION OF MULTIMERIC PROTEINS, AND RELATED
; FILE REFERENCE: 38814 351B
; CURRENT APPLICATION NUMBER: US/10/032,201B
; CURRENT FILING DATE: 2001-12-19
; NUMBER OF SEQ ID NOS: 313
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 204
; LENGTH: 312
; TYPE: PRT
; ORGANISM: Chlamydia trachomatis
US-10-032-201B-204

Query Match      70.2%; Score 33; DB 12; Length 312;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3 EGPFSG 8
DB      33 EGPFSG 38

RESULT 8
US-10-212-499-41
; Sequence 41, Application US/10212499
; Publication No. US20030135036A1
; GENERAL INFORMATION:
; APPLICANT: Lanza, Francois
; APPLICANT: Phillips, David R.
; APPLICANT: Cazenave, Jean-Pierre
; TITLE OF INVENTION: Platelet Glycoprotein V Gene and Uses
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morgan Lewis & Bockius LLP
; STREET: 1800 M St., NW
; CITY: Washington
; STATE: DC
; COUNTRY: US

; TITLE OF INVENTION: METHODS FOR THE PRODUCTION OF MULTIMERIC PROTEINS, AND RELATED
; FILE REFERENCE: 38814 351B
; CURRENT APPLICATION NUMBER: US/10/032,201B
; CURRENT FILING DATE: 2001-12-19
; NUMBER OF SEQ ID NOS: 313
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 202
; LENGTH: 312
; TYPE: PRT
; ORGANISM: Chlamydia muridarum
US-10-032-201B-202

Query Match      70.2%; Score 33; DB 12; Length 312;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3 EGPFSG 8
DB      33 EGPFSG 38

RESULT 9
US-09-828-326-20
; Sequence 20, Application US/09828326
; Publication No. US2003005952A1
; GENERAL INFORMATION:
; APPLICANT: Chait, Brian T.
; APPLICANT: Beavis, Ronald
; APPLICANT: Wang, Rong
; APPLICANT: Kent, Stephen B.H.
; TITLE OF INVENTION: Method and Product for the Sequence
; DETERMINATION OF PEPTIDES USING A MASS SPECTROMETER
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Crew LLP
```

; GENERAL INFORMATION:
; APPLICANT: DAVIS, Stacey
; TITLE OF INVENTION: METHOD AND COMPOSITIONS FOR INHIBITING THROMBIN-INDUCED COAGULATION
; FILE REFERENCE: P07201US01/BAS
; CURRENT APPLICATION NUMBER: US/10/142,935
; CURRENT FILING DATE: 2002-05-13
; PRIOR APPLICATION NUMBER: US 60/290,072
; PRIOR FILING DATE: 2001-05-13
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-10-142-935-5

Query Match 68.1%; Score 32; DB 15; Length 20;
Best Local Similarity 85.7%; Pred. No. 13;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 3 EGFESGR 9
Db 8 EGFESAR 14

RESULT 14
US-10-142-935-6
; Sequence 6, Application US/10142935
; Publication No. US2003004418A1
; GENERAL INFORMATION:
; APPLICANT: DAVIS, Stacey
; TITLE OF INVENTION: METHOD AND COMPOSITIONS FOR INHIBITING THROMBIN-INDUCED COAGULATION
; FILE REFERENCE: P07201US01/BAS
; CURRENT APPLICATION NUMBER: US/10/142,935
; CURRENT FILING DATE: 2002-05-13
; PRIOR APPLICATION NUMBER: US 60/290,072
; PRIOR FILING DATE: 2001-05-13
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-10-142-935-6

Query Match 68.1%; Score 32; DB 15; Length 20;
Best Local Similarity 85.7%; Pred. No. 13;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 3 EGFESGR 9
Db 3 EGFESAR 9

RESULT 15
US-10-142-935-4
; Sequence 4, Application US/10142935
; Publication No. US2003004418A1
; GENERAL INFORMATION:
; APPLICANT: DAVIS, Stacey
; TITLE OF INVENTION: METHOD AND COMPOSITIONS FOR INHIBITING THROMBIN-INDUCED COAGULATION
; FILE REFERENCE: P07201US01/BAS
; CURRENT APPLICATION NUMBER: US/10/142,935
; CURRENT FILING DATE: 2002-05-13
; PRIOR APPLICATION NUMBER: US 60/290,072
; PRIOR FILING DATE: 2001-05-13
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 25

; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-10-142-935-4
Query Match 68.1%; Score 32; DB 15; Length 25;
Best Local Similarity 85.7%; Pred. No. 17;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 3 EGFESGR 9
Db 8 EGFESAR 14

Search completed: December 11, 2003, 18:38:02
Job time : 49.3333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 18:19:54 ; Search time 8.3333 Seconds
(without alignments)
103.862 Million cell updates/sec

Title: US-10-014-658-4

Perfect score: 47

Sequence: 1 STEGFFSGR 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 76:*

1: Pirl:*

2: Pirl2:*

3: Pirl3:*

4: Pirl4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	80.9	700	2 F86622	Cr356 hypothetical protein
2	38	80.9	700	2 F72001	conserved hypothetical
3	36	76.6	441	2 D83555	hypothetical prote
4	36	76.6	518	2 JC7387	testis-abundant fi
5	36	76.6	651	2 AD3057	glycogen debranchi
6	36	76.6	651	2 B98229	glycogen debranchi
7	35	74.5	323	2 AH0608	NADH oxidoreductas
8	35	74.5	1302	2 T23236	hypothetical prote
9	34	72.3	97	2 E83584	hypothetical prote
10	34	72.3	322	2 F90748	NADH oxidoreductas
11	34	72.3	322	2 B85599	probable enzyme Z1
12	34	72.3	322	2 H54835	hypothetical prote
13	34	72.3	383	2 T26902	hypothetical prote
14	34	72.3	1882	1 GNVVIR	genome polyprotein
15	33	70.2	311	2 C72030	thioredoxin reduct
16	33	70.2	312	2 C72030	thioredoxin reduct
17	33	70.2	312	2 C81710	thioredoxin reduct
18	33	70.2	317	2 AD1542	oxidoreductases ho
19	33	70.2	317	2 AP1184	oxidoreductases ho
20	33	70.2	351	2 B11586	probable thioredox
21	33	70.2	390	2 D96723	hypothetical prote
22	33	70.2	452	1 D69810	phosphotransferase
23	33	70.2	627	2 H96951	fusion, PTS system
24	33	70.2	759	2 F70539	probable 5-methylt
25	33	70.2	959	2 T38247	probable alanyl-tr
26	32	68.1	110	2 T00622	hypothetical prote
27	32	68.1	143	2 D11009	hypothetical prote
28	32	68.1	223	2 B11254	hypothetical prote
29	32	68.1	253	2 AE3511	homospermidine syn

30 32 68.1 307 2 F84162
31 32 68.1 344 2 C69453
32 32 68.1 361 2 F97068
33 32 68.1 368 2 F90152
34 32 68.1 379 1 S55421
35 32 68.1 413 2 C33225
36 32 68.1 448 2 B85077
37 32 68.1 468 2 JC4285
38 32 68.1 491 1 FGHUB
39 32 68.1 497 2 T16900
40 32 68.1 589 2 T34878
41 32 68.1 778 2 D95912
42 32 68.1 824 2 F72408
43 31 66.0 82 2 S06772
44 31 66.0 181 1 RKRVS
45 31 66.0 181 1 RKRPS

ALIGNMENTS

RESULT 1

F86622

Cr356 hypothetical protein [imported] - Chlamydoiphila pneumoniae (strain J138)
C:Species: Chlamydoiphila pneumoniae, Chlamydia pneumoniae

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001

C:Accession: F86622

R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Ouchi, K.; Shiba, T.; I

Nucleic Acids Res. 28, 2311-2314, 2000

A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.

A:Reference number: A86491; MUID:20330349; PMID:10871362

A:Accession: F86622

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-700 <STO>

A:Cross-references: GB:BA000008; NID:98979430; PIDN:BA999264.1; GSPDB:GN00142

A:Experimental source: strain J138

C:Genetics:

A:Gene: CFJ1057

C:Superfamily: conserved hypothetical protein yyaL

Query Match 80.9%; Score 38; DB 2; Length 700;
Best Local Similarity 77.8%; Pred. No. 11;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 STEGFFSGR 9
| | | | |
Db 367 SREGFFNGR 375

RESULT 2

F72001

conserved hypothetical protein CP0793 [imported] - Chlamydoiphila pneumoniae (strains CW
C:Species: Chlamydoiphila pneumoniae, Chlamydia pneumoniae

C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000

C:Accession: F72001; E81536

R:Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.

Nature Genet. 21, 385-389, 1999

A:Title: Comparative Genomes of Chlamydia pneumoniae and C. trachomatis.

A:Reference number: A72000; MUID:99206606; PMID:10192388

A:Accession: F72001

A:Molecule type: DNA

A:Residues: 1-700 <ARN>

A:Cross-references: GB:AE001686; GB:AE001363; NID:94377389; PIDN:AA019194.1; PID:943773

A:Experimental source: strain CWL029

R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,

, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg

Nucleic Acids Res. 28, 1397-1406, 2000

A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.

A:Reference number: A81500; MUID:20150255; PMID:10684935

A:Accession: E81536

A:Molecule type: DNA

A:Residues: 1-700 <REA>

C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C;Accession: AH0608
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Conington, P.; Cronin, A.; Davies, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A;Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serov
A;Reference number: AB0502; MUID:21534947; PMID:11677608
A;Accession: AH0608
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-323 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD05338.1; PID:G16502102; GSPDB:GN00176
C;Genetics:
C;Gene: hcr
C;Superfamily: phthalate dioxygenase reductase; cytochrome-b5 reductase homology; ferred
C;Keywords: oxidoreductase

Query Match 74.5%; Score 35; DB 2; Length 323;
Best Local Similarity 66.7%; Pred. No. 21;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 STGFFSGR 9
:|||||:
Db 174 ATEGFIAGR 182

RESULT 8
T23236
hypothetical protein K02C4.3 - *Caenorhabditis elegans*
C;Species: *Caenorhabditis elegans*
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T23236
R;Lightning, J.
submitted to the EMBL Data Library, January 1995
A;Reference number: Z19713
A;Accession: T23236
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1302 <WIL>
A;Cross-references: EMBL:Z47811; PIDN:CAA87786.1; GSPDB:GN00020; CESP:K02C4.3
A;Experimental source: clone K02C4
C;Genetics:
A;Gene: CESP:K02C4.3
A;Map position: 2
A;Introns: 10/1; 31/3; 84/3; 279/3; 464/1; 745/1; 957/2; 978/3; 1090/2; 1201/3; 1270/3

Query Match 74.5%; Score 35; DB 2; Length 1302;
Best Local Similarity 75.0%; Pred. No. 88;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 TEGFFSGR 9
:|||||:
Db 669 TEGFFPGK 676

RESULT 9
E83594
hypothetical protein PA0490 [imported] - *Pseudomonas aeruginosa* (strain PA01)
C;Species: *Pseudomonas aeruginosa*
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: E83584
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic patho
A;Reference number: AB2950; MUID:20437337; PMID:10984043
A;Accession: E83584
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-97 <STO>

A;Cross-references: GB:AE004486; GB:AE004091; MID:99946345; PIDN:AA03879.1; GSPDB:GN00
A;Experimental source: strain PA01
C;Genetics:
A;Gene: PA0490.

Query Match 72.3%; Score 34; DB 2; Length 97;
Best Local Similarity 85.7%; Pred. No. 10;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TEGFFSG 8
:|||||:
Db 3 TEGFFDG 9

RESULT 10
F90748
NADH oxidoreductase for HCP EC0958 [imported] - *Escherichia coli* (strain O157:H7, subs
C;Species: *Escherichia coli*
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 24-Aug-2001
C;Accession: F90748
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and gen
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: F90748
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-322 <HAY>
A;Cross-references: PIDN:BAE34391.1; PID:G13360417; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RMD 050952
C;Genetics:
A;Gene: EC0958
C;Superfamily: phthalate dioxygenase reductase; cytochrome-b5 reductase homology; ferre

Query Match 72.3%; Score 34; DB 2; Length 322;
Best Local Similarity 75.0%; Pred. No. 34;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 TEGFFSGR 9
:|||||:
Db 174 TEGFIAGR 181

RESULT 11
B85599
probable enzyme Z1106 [imported] - *Escherichia coli* (strain O157:H7, substrain EDL933)
C;Species: *Escherichia coli*
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: B85599
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhe
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lam, A.; Dimalanta, E.; Potamousis, K.; Apodaca
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: B85599
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-322 <STO>
A;Cross-references: GB:AE005174; MID:G12513899; PIDN:AA055254.1; GSPDB:GN00145; UWGP:Z1
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: Z1106
C;Superfamily: phthalate dioxygenase reductase; cytochrome-b5 reductase homology; ferre

Query Match 72.3%; Score 34; DB 2; Length 322;
Best Local Similarity 75.0%; Pred. No. 34;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 TEGFFSGR 9
:|||||:
Db 174 TEGFIAGR 181

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 16:54:27 ; Search time 4.66667 Seconds
(without alignments)
90.694 Million cell updates/sec

Title: US-10-014-658-4
Perfect score: 47
Sequence: 1 STEGFFSGR 9

Scoring table:
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36	76.6	488	1 RN23 MOUSE	Q9esn2 mus musculus
2	36	76.6	518	1 RN23 HUMAN	Q9hcn9 homo sapien
3	35	74.5	1302	1 UBPY CAEL	Q09931 caenorhabdi
4	34	72.3	322	1 HCR ECOLI	P75824 escherichia
5	34	72.3	1882	1 POL2 TRSVR	P25247 tomato ring
6	33	70.2	92	1 VC02 SPVKA	P32230 swinepox vi
7	33	70.2	311	1 TEXB CHLPN	Q3z8m4 chlamydia p
8	33	70.2	312	1 TRXB CHLMU	Q9pkr7 chlamydia m
9	33	70.2	312	1 TRXB CHLTR	O84101 chlamydia t
10	33	70.2	759	1 METE MYCTU	O06584 mycobacteri
11	33	70.2	959	1 SYAC SCHPO	O13914 schizosacch
12	32	68.1	110	1 Y103 ARATH	O80594 arabidopsis
13	32	68.1	223	1 YAO0 TRBPA	O83965 treponema p
14	32	68.1	379	1 ACDA BACSU	P45857 bacillus su
15	32	68.1	468	1 PALP PICPA	P52231 picchia past
16	32	68.1	491	1 FIBB HUMAN	P02675 homo sapien
17	32	68.1	629	1 SYR PYRFU	Q8u149 pyrococcus
18	32	68.1	824	1 SYL THERM	Q9wy15 thermotoga
19	32	68.1	887	1 K20A MOUSE	P37323 mus musculu
20	32	68.1	890	1 K20A HUMAN	O95235 homo sapien
21	31	66.0	82	1 RBS SIGNAL	P13951 sinapis alb
22	31	66.0	181	1 RBS1 BRANA	P05346 brassica na
23	31	66.0	181	1 RBS2 BRANA	P27985 brassica na
24	31	66.0	181	1 RBS RAPSA	P08135 raphrassica na
25	31	66.0	188	1 YAVD RHISN	Q33212 rhizobium s
26	31	66.0	198	1 C4BB BOVIN	Q28066 bos taurus
27	31	66.0	200	1 YE01 MYCTU	P71669 mycobacteri
28	31	66.0	349	1 CRMB CAMPS	O8uya7 camelpox vi
29	31	66.0	349	1 CRMB VAVV	P34015 variola vir
30	31	66.0	351	1 CRMB COMPX	O73559 cowpox viru
31	31	66.0	674	1 YFE2 YEAST	P43560 saccharomyc
32	31	66.0	731	3 X1LS SULSO	Q9p999 sulfolobus
33	31	66.0	734	1 GLGB_AGRTU	P52979 agrobacteri

RESULT 1					
RN23_MOUSE					
ID	RN23_MOUSE	STANDARD;	PRT;	488 AA.	
AC	Q9ESN2;				
DT	16-OCT-2001 (Rel. 40, Created)				
DT	16-OCT-2001 (Rel. 40, Last sequence update)				
DT	28-FEB-2003 (Rel. 41, Last annotation update)				
DE	RING finger protein 23 (Testis-abundant finger protein) (Tripartite				
DE	motif-containing protein 39).				
GN	TRIM39 OR RNFP23 OR YFP.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Testis;				
RX	MEDLINE=20462913; PubMed=11006080;				
RA	Orimo A., Yamagishi T., Tomimaga N., Yamauchi Y., Hishinuma T.,				
RA	Okada K., Suzuki M., Sato M., Nogi Y., Suzuki H., Inoue S.,				
RA	Yoshimura K., Shimizu Y., Muramatsu M.;				
RT	"Molecular cloning of testis-abundant finger protein/ring finger				
RT	protein 23 (RNFP23), a novel RING-B box-coiled coil-B30.2 protein on				
RT	the class I region of the human MHC";				
RL	Biochem. Biophys. Res. Commun. 276:45-51(2000).				
CC	-!- SIMILARITY: Contains 1 RING-type zinc finger.				
CC	-!- SIMILARITY: Contains 1 B box-type zinc finger.				
CC	-!- SIMILARITY: Contains 1 SPRY domain.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
CC	EMBL; AB046382; BAB16375.1; --				
DR	HSSP; P15919; IRMD.				
DR	MGD; MGI:1890859; Trim39.				
DR	InterPro; IPR001870; B302.				
DR	InterPro; IPR006577; SPRY.				
DR	InterPro; IPR003877; SPRY receptor.				
DR	InterPro; IPR000315; Znf Bbox.				
DR	InterPro; IPR001841; Znf_ring.				
DR	Pfam; PF00822; SPRY; 1.				
DR	Pfam; PF00843; zf-B_box; 1.				
DR	Pfam; PF00097; zf-C3HC4; 1.				
DR	SMART; SM00336; BBOX; 1.				
DR	SMART; SM00589; PRY; 1.				
DR	SMART; SM00184; RING; 1.				
DR	SMART; SM00449; SPRY; 1.				
DR	PROSITE; PS50119; ZF_BOX; 1.				
DR	PROSITE; PS00518; ZF_RING_1; 1.				
DR	PROSITE; PS50089; ZF_RING_2; 1.				
KW	Zinc-finger; Coiled coil.				

Q27319 homarus ame
Q9hzn4 pseudomonas
Q9zmv1 helicobacte
P56045 helicobacte
Q9piq0 campylobact
Q04569 lelystad vi
P72873 synchocyst
Q06562 cochiobolu
Q43097 thermomyces
Q49823 mycobacteri
Q11169 mycobacteri
Q10108 schizosacch

ALIGNMENTS


```

CC -!- SIMILARITY: Belongs to peptidase family C19.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z47811; CAA87786.1; -
CC PIR; T23236; T23236.
CC MEROPS; C19; UFW; -
CC WormPep; K02C4.3; CB01599.
CC InterPro; IPR001394; UCH-2.
CC Pfam; PF00443; UCH; 1.
CC PROSITE; PS00972; UCH_2_1; 1.
CC PROSITE; PS00973; UCH_2_2; 1.
CC PROSITE; PS00974; UCH_2_3; 1.
CC PROSITE; PS00975; UCH_2_4; 1.
CC Hypothetical protein; UBL conjugation pathway; Hydrolase;
KW Thiol protease.
FT ACT SITE 208 BY SIMILARITY.
FT ACT SITE 730 BY SIMILARITY.
FT ACT SITE 738 BY SIMILARITY.
FT ACT SITE 738 BY SIMILARITY.
SQ SEQUENCE 1302 AA; 147790 MW; 3097CE581963D033 CRC64;

Query Match 74.5%; Score 35; DB 1; Length 1302;
Best Local Similarity 75.0%; Pred. No. 40;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 TEGFFSGR 9
Db 669 TEGFFPGK 676

RESULT 4
ID -HCR_ECOLI STANDARD; PRT; 322 AA.
AC P75824;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE NADH oxidoreductase hcr (EC 1.-.-.-).
HCR CR B0872.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RA "The complete genome sequence of Escherichia coli K-12."
RT Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikenoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map."
RL DNA Res. 3:137-155(1996).
RN [3]
RP CHARACTERIZATION.

RX MEDLINE=20117674; PubMed=10651802;
RA van den Berg W.A.M., Hagen W.R., van Dongen W.M.A.M.;
RT "The hybrid-cluster protein ('prismane protein') from Escherichia
RT coli. Characterization of the hybrid-cluster protein, redox properties
RT of the [2Fe-2S] and [4Fe-2S-2O] clusters and identification of an
RT associated NADH oxidoreductase containing FAD and 2Fe-2S."
RL Bur. J. Biochem. 267:666-676(2000).
CC -!- FUNCTION: NADH OXIDOREDUCTASE ACTING IN CONCERT WITH HCP.
CC -!- COFACTOR: FAD AND A 2FE-2S CLUSTER.
CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE FAD
CC FLAVOPROTEIN OXIDOREDUCTASES FAMILY.
CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE 2FE2S PLANT-
CC TYPE FERREDOXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000189; AAC73959.1; -
CC PIR; H64825; BAA35586.1; ALT_INIT.
CC HSP; P00235; IFR.
CC EcoGene; EG1361; hcr.
CC InterPro; IPR001834; Cyt_B5_reductase.
CC InterPro; IPR001041; Ferredoxin.
CC InterPro; IPR001433; Oxid_FAD/NAD(P).
CC InterPro; IPR000951; Phdiox_reductase.
CC InterPro; IPR001221; Phe_hydroxylase.
CC Pfam; PF00970; FAD_binding_6; 1.
CC Pfam; PF00111; fer2; 1.
CC Pfam; PF00175; NAD_binding_1; 1.
CC PRINTS; PR00409; PHDIOXRDTASE.
CC PRINTS; PR00410; PHEHYDRXLASE.
KW Oxidoreductase; Flavoprotein; FAD; NAD; Metal-binding; Iron-sulfur;
KW Iron; 2Fe-2S; Electron transport; Complete proteome.
FT DOMAIN 111 213 OXIDOREDUCTASE.
FT DOMAIN 239 322 FERREDOXIN.
FT METAL 273 273 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 278 278 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 281 281 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 311 311 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
SQ SEQUENCE 322 AA; 35740 MW; 5138026DCBBD6A9C CRC64;

Query Match 72.3%; Score 34; DB 1; Length 322;
Best Local Similarity 75.0%; Pred. No. 15;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 TEGFFSGR 9
Db 174 TEGFIAGR 181

RESULT 5
ID -POL2_TRSVR STANDARD; PRT; 1882 AA.
AC P25247;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE RNA2 polyprotein (207 kDa protein) [Contains: Coat protein].
OS Tomato ringspot virus (isolate raspberry) (TOMRSV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Comoviridae;
OC Nepovirus.
OX NCBI_TaxID=12281;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91311402; PubMed=1856689;
RA Rott M.E., Tremaine J.H., Rochon D.M.;
RT "Nucleotide sequence of tomato ringspot virus RNA-2."

```

```
DR EMBL; AP002546; BAA98524.1; --
DR PIR; B86530; B86530.
DR PIR; C81710; C81710.
DR HSSP; Q39243; 1VDC.
DR PHCI-2DPAGE; Q928M4; --
DR TIGR; CP0444; --
DR InterPro; IPR000759; Adrmdx_reductase.
DR InterPro; IPR001327; FAD_pyr_redux.
DR InterPro; IPR00103; Pyridine_redux.2.
DR InterPro; IPR005982; Thio-redox_redux.
DR Pfam; PF00070; Pyr_redux.1.
DR PRINTS; PR00419; ADXRDTASE.
DR PRINTS; PR00368; FADPNR.
DR PRINTS; PR00469; PNRDRTASEII.
DR TIGRFAMS; TIGR01292; TRX_reduct; 1.
DR PROSITE; PS00573; PYRIDINE REDOX 2; 1.
KW Redox-active center; Oxidoreductase; NADP; Flavoprotein; FAD;
KW Complete proteome.
FT NP_BIND 4 19 FAD (ADP PART) (PROBABLE).
FT DISULFID 138 141 REDOX-ACTIVE (BY SIMILARITY).
FT NP_BIND 152 166 NAD(P) (BY SIMILARITY).
FT NP_BIND 273 283 FAD (FLAVIN PART) (BY SIMILARITY).
SQ SEQUENCE 311 AA; 33559 MW; 352CE58785C2B27 CRC64;

Query Match 70.2%; Score 33; DB 1; Length 311;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 EGFPFG 8
DB 33 EGFPFG 38

RESULT 8
TRXB_CHLMU STANDARD; PRT; 312 AA.
AC Q9PKT7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Thio-redoxin reductase [EC 1.8.1.9] (TRXR).
GN TRXB OR TC0375.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MoPn / Ni99;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Baas S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -!- CATALYTIC ACTIVITY: Thio-redoxin + NADP(+) = thio-redoxin disulfide
CC + NADPH.
CC -!- COFACTOR: FAD.
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- MISCELLANEOUS: THE ACTIVE SITE IS A REDOX-ACTIVE DISULFIDE BOND.
CC -!- SIMILARITY: BELONGS TO THE PYRIDINE NUCLEOTIDE-DISULFIDE
CC OXIDOREDUCTASES CLASS-II.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

CC EMBL; AP002546; BAA98524.1; --
CC PIR; B86530; B86530.
CC PIR; C81710; C81710.
CC HSSP; Q39243; 1VDC.
CC PHCI-2DPAGE; Q928M4; --
CC TIGR; CP0444; --
CC InterPro; IPR000759; Adrmdx_reductase.
CC InterPro; IPR001327; FAD_pyr_redux.
CC InterPro; IPR00103; Pyridine_redux.2.
CC InterPro; IPR005982; Thio-redox_redux.
CC Pfam; PF00070; Pyr_redux.1.
CC PRINTS; PR00419; ADXRDTASE.
CC PRINTS; PR00368; FADPNR.
CC PRINTS; PR00469; PNRDRTASEII.
CC TIGRFAMS; TIGR01292; TRX_reduct; 1.
CC PROSITE; PS00573; PYRIDINE REDOX 2; 1.
KW Redox-active center; Oxidoreductase; NADP; Flavoprotein; FAD;
KW Complete proteome.
CC NP_BIND 4 19 FAD (ADP PART) (PROBABLE).
CC DISULFID 138 141 REDOX-ACTIVE (BY SIMILARITY).
CC NP_BIND 152 166 NAD(P) (BY SIMILARITY).
CC NP_BIND 273 283 FAD (FLAVIN PART) (BY SIMILARITY).
CC SEQUENCE 311 AA; 33559 MW; 352CE58785C2B27 CRC64;

Query Match 70.2%; Score 33; DB 1; Length 312;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 EGFPFG 8
DB 33 EGFPFG 38

RESULT 9
TRXB_CHLTR STANDARD; PRT; 312 AA.
AC O84101;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Thio-redoxin reductase [EC 1.8.1.9] (TRXR).
GN TRXB OR CT0099.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UW-3/Cx;
RX MEDLINE=99000809; PubMed=97841136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.D., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis.";
RL Science 282:754-759(1998).
CC -!- CATALYTIC ACTIVITY: Thio-redoxin + NADP(+) = thio-redoxin disulfide
CC + NADPH.
CC -!- COFACTOR: FAD.
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- MISCELLANEOUS: THE ACTIVE SITE IS A REDOX-ACTIVE DISULFIDE BOND.
CC -!- SIMILARITY: BELONGS TO THE PYRIDINE NUCLEOTIDE-DISULFIDE
CC OXIDOREDUCTASES CLASS-II.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

CC EMBL; AB001284; AAC67690.1; ALT_INIT.
CC HSSP; Q39243; 1VDC.
```

RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreno S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forburg S.L.,
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakowski G.V., Usery D., Barrell B.G., Nurse P.,
 RT "The genome sequence of Schizosaccharomyces pombe";
 RL Nature 415:971-980(2002).
 CC -1- CATALYTIC ACTIVITY: ATP + L-alanine + tRNA(Ala) = AMP +
 CC di-phosphate + L-alanyl-tRNA(Ala).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC
 DR EMBL; Z98559; CAB11162.1; -
 DR PIR; T38247; T38247.
 DR GeneDB SPombe; SPAC23C11.09; -
 DR InterPro; IPR003156; DHHAL.
 DR InterPro; IPR002318; tRNA-synt_2c.
 DR InterPro; IPR006193; tRNA_synt_Ala.
 DR Pfam; PF02272; DHHAL; 1. 24
 DR Pfam; PF01411; tRNA-synt_2c; 1. 24
 DR PRINTS; PRO0380; TRNASYNTHALA.
 DR TIGRFAMs; TIGR00344; alas; 1.
 DR PROSITE; PS0860; AA TRNA LIGASE II ALA; 1.
 KW Aminoacyl-tRNA synthetase; Protein Biosynthesis; Ligase; ATP-binding.
 SQ SEQUENCE 959 AA; 107376 MW; 9F2E4A725D866F86 CRC64;

Query Match 70.2%; Score 33; DB 1; Length 959;
 Best Local Similarity 87.5%; Pred. No. 75;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 STEGFPSG 8
 |||||
 DB 520 STEGFNSG 527

RESULT 12
 Y103 ARATH
 ID Y103 ARATH STANDARD; PRT; 110 AA.
 AC O80594;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein At1g10030.
 GN AT1G10030 OR T2711.5
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsidae.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=21016719; PubMed=11130712;
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
 RA White O., Alonso J., Altieri H., Araujo R., Bowman C.L., Brooks S.Y.,
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
 RA Dunn P., Egtu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
 RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,

RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
 RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
 RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
 RA Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
 RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
 RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
 RT thaliana";
 RL Nature 408:816-820(2000).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: BELONGS TO THE UPF0143 FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC
 DR EMBL; AC004122; AAC34343.1; -
 DR PIR; T00622; T00622. UPF0143.
 DR InterPro; IPR005352; UPF0143.
 DR Pfam; PF03694; Erg28; 1.
 DR ProDom; PD089646; UPF0143; 1.
 KW Hypothetical protein; Transmembrane.
 FT TRANSMEM 4 24 POTENTIAL.
 FT TRANSMEM 46 56 POTENTIAL.
 FT TRANSMEM 71 91 POTENTIAL.
 SQ SEQUENCE 110 AA; 12550 MW; E74593110F8669BB CRC64;

Query Match 68.1%; Score 32; DB 1; Length 110;
 Best Local Similarity 75.0%; Pred. No. 13;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 STEGFPSG 8
 |||||
 DB 101 STVGFFAG 108

RESULT 13
 YA00 TREPA
 ID YA00 TREPA STANDARD; PRT; 223 AA.
 AC O83955;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein TP1000.
 GN TP1000.
 OS Treponema pallidum.
 OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
 OC NCBI_TaxID=160;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Nichols;
 RX MEDLINE=98332770; PubMed=9655876;
 RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,
 RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
 RA Sodergren B., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
 RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
 RA McDonald L., Artach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
 RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
 RA Venter J.C.;
 RT "Complete genome sequence of Treponema pallidum, the syphilis
 RT spirochete";
 RL Science 281:375-388(1998).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; U28658; AAA85503.1; -.
DR PIR; JC4285; JC4285.
DR HSP; P34755; IQFX.
DR InterPro; IPR000560; HisAc phsphtse.
DR Pfam; PF00328; acid phosphat; 1.
DR PROSITE; PS00616; HIS_ACID_PHOSPHAT_1; 1.
DR PROSITE; PS00778; HIS_ACID_PHOSPHAT_2; 1.
KW Hydrolase; Glycoprotein; Signal; Multigene family.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 468 ACID PHOSPHATASE PHO1.
FT ACT_SITE 84 84 NUCLEOPHILIC ACCEPTOR (BY SIMILARITY).
FT ACT_SITE 345 345 PROTON DONOR (BY SIMILARITY).
FT CARBOHYD 163 163 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 256 256 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 321 321 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 360 360 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 453 453 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 468 AA; 52690 MW; AE555B8E27718C2C CRC64;

Query Match 68.1%; Score 32; DB 1; Length 468;
Best Local Similarity 62.5%; Pred. No. 57;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 STEGFFSG 8
Db 136 TTKGFYSG 143

Search completed: December 11, 2003, 18:20:56
Job time : 7.66667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 18:03:30 ; Search time 20.3333 Seconds
(without alignments)
114.220 Million cell updates/sec

Title: US-10-014-658-4

Perfect score: 47

Sequence: 1 STEGFFSGR 9

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SP_TREMBL 23.3*

2: SP_Archea:

3: SP_Bacteria:

4: SP_Fungi:

5: SP_Human:

6: SP_Invertebrate:

7: SP_Mammal:

8: SP_MHC:

9: SP_Organelle:

10: SP_Phage:

11: SP_Plant:

12: SP_Rodent:

13: SP_Virus:

14: SP_Vertebrate:

15: SP_Unclassified:

16: SP_Virus:

17: SP_Archeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	80.9	700	Q926K2	Q926K2 chlamydia p
2	36	76.6	200	O55482	O55482 porcine rep
3	36	76.6	200	O41187	O41187 porcine rep
4	36	76.6	200	Q9J7K2	Q9J7K2 porcine rep
5	36	76.6	200	O990Q8	O990Q8 porcine rep
6	36	76.6	200	O91CC5	O91CC5 porcine rep
7	36	76.6	200	O55486	O55486 porcine rep
8	36	76.6	200	O55490	O55490 porcine rep
9	36	76.6	200	O41195	O41195 porcine rep
10	36	76.6	200	O41195	O41195 porcine rep
11	36	76.6	200	O41195	O41195 porcine rep
12	36	76.6	441	Q9J7J5	Q9J7J5 porcine rep
13	36	76.6	488	Q915L2	Q915L2 pseudomonas
14	36	76.6	496	Q8K0F7	Q8K0F7 mus musculus
15	36	76.6	651	Q8U8L7	Q8U8L7 agrobacteri
16	35	74.5	323	Q8ZQE9	Q8ZQE9 salmonella

17	35	74.5	323	16	Q8Z830	Q8Z830 salmonella
18	35	74.5	400	16	O8E282	O8E282 leptospira
19	34	72.3	97	16	Q91528	Q91528 pseudomonas
20	34	72.3	201	12	O8U5H5	O8U5H5 porcine rep
21	34	72.3	201	12	O8JY7	O8JY7 porcine rep
22	34	72.3	284	16	O8DH47	O8DH47 synchococc
23	34	72.3	322	16	Q8X6L2	Q8X6L2 escherichia
24	34	72.3	324	16	O8FUE1	O8FUE1 escherichia
25	34	72.3	383	5	O45938	O45938 caenorhabdi
26	34	72.3	562	12	O3QRA4	O3QRA4 tomato ring
27	34	72.3	562	12	O3QRA3	O3QRA3 tomato ring
28	34	72.3	562	12	O3QRA5	O3QRA5 tomato ring
29	34	72.3	564	12	O3YPA2	O3YPA2 tomato ring
30	33	70.2	192	2	Q8S5B9	Q8S5B9 desulfovibr
31	33	70.2	200	12	Q990Q4	Q990Q4 porcine rep
32	33	70.2	200	12	Q9J7B6	Q9J7B6 porcine rep
33	33	70.2	200	12	O41182	O41182 porcine rep
34	33	70.2	200	12	O41183	O41183 porcine rep
35	33	70.2	200	12	Q9WHH1	Q9WHH1 porcine rep
36	33	70.2	200	12	Q9J7L4	Q9J7L4 porcine rep
37	33	70.2	200	12	Q9J7K3	Q9J7K3 porcine rep
38	33	70.2	200	12	O88701	O88701 porcine rep
39	33	70.2	200	12	O41191	O41191 porcine rep
40	33	70.2	200	12	O8QQW4	O8QQW4 porcine rep
41	33	70.2	200	12	Q9J7L3	Q9J7L3 porcine rep
42	33	70.2	200	12	Q9WGE4	Q9WGE4 porcine rep
43	33	70.2	200	12	Q990Q9	Q990Q9 porcine rep
44	33	70.2	200	12	O9YLS8	O9YLS8 porcine rep
45	33	70.2	200	12	O56258	O56258 porcine rep

ALIGNMENTS

RESULT 1

Q926K2 PRELIMINARY; PRT; 700 AA.

ID Q926K2
AC Q926K2
DT 01-MAY-1999 (TREMELrel. 10, Created)
DT 01-MAY-1999 (TREMELrel. 10, Last sequence update)
DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)
DE CT356 hypothetical protein.
GN CPN1057 OR CPJ1057 OR CP0793.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RX MEDLINE=9206606; PubMed=10192388;
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
White O., Hickey E.K., Petersen J., Utterback T., Berry K., Bass S.,
Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Cuchi K.,
Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138

RA Key K.F., Haqshenas G., Guenette D.K., Swenson S.L., Toth T.E.,
RA Meng X.J.;
RT "Genetic Characterization of the Major Envelope Gene of Acute Porcine
RT Reproductive and Respiratory Syndrome Virus Isolates.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF339494; AAK25804.1;
DR InterPro; IPR001332; Arteri_glycop.
DR InterPro; IPR003239; Porcine_RR_virus.
DR Pfam; PF00951; Arteri_glycop; 1.
DR ProDom; PD001151; Porcine_RR_virus; 1.
SQ SEQUENCE 200 AA; 22379 MW; 1A06BF83D3717D09 CRC64;

Query Match 76.6%; Score 36; DB 12; Length 200;
Best Local Similarity 77.8%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 STEGFFSGR 9
||| ||| |||
Db 97 STAGFFHGR 105

RESULT 6

ID Q9ICCS PRELIMINARY; PRT; 200 AA.
AC Q9ICCS;

DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE Envelope glycoprotein.
OS Porcine reproductive and respiratory syndrome virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Arteriviridae; Arterivirus.
OX NCBI_TaxID=28344;
RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=PRRSV21, PRESV17, and PRESV20;
RX MEDLINE=20109045; PubMed=10640555;
RA Goldberg T.L., Hahn E.C., Weigel R.M., Scherba G.;
RT "Genetic, geographical and temporal variation of porcine reproductive
RT and respiratory syndrome virus in Illinois.";
RL J. Gen. Virol. 81:171-179 (2000).

DR EMBL; AF176443; AAF36249.1;
DR EMBL; AF176439; AAF36245.1;
DR EMBL; AF176442; AAF36248.1;
DR InterPro; IPR001332; Arteri_glycop.
DR InterPro; IPR003239; Porcine_RR_virus.
DR Pfam; PF00951; Arteri_glycop; 1.
DR ProDom; PD001151; Porcine_RR_virus; 1.
SQ SEQUENCE 200 AA; 22404 MW; BD1223E0FB7D0CB CRC64;

Query Match 76.6%; Score 36; DB 12; Length 200;
Best Local Similarity 77.8%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 STEGFFSGR 9
||| ||| |||
Db 97 STAGFFHGR 105

RESULT 7

ID O55486 PRELIMINARY; PRT; 200 AA.
AC O55486;

DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Gp5.
OS Porcine reproductive and respiratory syndrome virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Arteriviridae; Arterivirus.
OX NCBI_TaxID=28344;
RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Danish DK5163-17;
RX MEDLINE=99003922; PubMed=9787654;
RA Madsen K.G., Hansen C.M., Madsen E.S., Strandbygaard B., Boetner A.,
RA Soerensen K.J.;
RT "Sequence analysis of porcine reproductive and respiratory syndrome
RT virus of the American type collected from Danish swine herds.";
RL Arch. Virol. 143:1683-1700 (1998).

DR EMBL; AJ223080; CAA11094.1;
DR InterPro; IPR001332; Arteri_glycop.
DR InterPro; IPR003239; Porcine_RR_virus.
DR Pfam; PF00951; Arteri_glycop; 1.
DR ProDom; PD001151; Porcine_RR_virus; 1.
SQ SEQUENCE 200 AA; 22407 MW; 99633E525748CA89 CRC64;

Query Match 76.6%; Score 36; DB 12; Length 200;
Best Local Similarity 77.8%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 STEGFFSGR 9
||| ||| |||
Db 97 STAGFFHGR 105

RESULT 8

ID O55490 PRELIMINARY; PRT; 200 AA.
AC O55490;

DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Gp5.
OS Porcine reproductive and respiratory syndrome virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Arteriviridae; Arterivirus.
OX NCBI_TaxID=28344;
RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Danish DK5163-23;
RX MEDLINE=99003922; PubMed=9787654;
RA Madsen K.G., Hansen C.M., Madsen E.S., Strandbygaard B., Boetner A.,
RA Soerensen K.J.;
RT "Sequence analysis of porcine reproductive and respiratory syndrome
RT virus of the American type collected from Danish swine herds.";
RL Arch. Virol. 143:1683-1700 (1998).

DR EMBL; AJ223081; CAA11100.1;
DR InterPro; IPR001332; Arteri_glycop.
DR InterPro; IPR003239; Porcine_RR_virus.
DR Pfam; PF00951; Arteri_glycop; 1.
DR ProDom; PD001151; Porcine_RR_virus; 1.
SQ SEQUENCE 200 AA; 22414 MW; 4BE85506A741DA49 CRC64;

Query Match 76.6%; Score 36; DB 12; Length 200;
Best Local Similarity 77.8%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 STEGFFSGR 9
||| ||| |||
Db 97 STAGFFHGR 105

RESULT 9

ID O41195 PRELIMINARY; PRT; 200 AA.
AC O41195;

DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Envelope protein.
GN ENV.
OS Porcine reproductive and respiratory syndrome virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Arteriviridae; Arterivirus.
OX NCBI_TaxID=28344;

RP SEQUENCE FROM N.A.

DE Similar to tripartite motif protein 39.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Strausberg R.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC034985; AAKH4985.1; -.
SQ SEQUENCE 488 AA; 56374 MW; 102AA8E5C8796A3E CRC64;

Query Match 76.6%; Score 36; DB 4; Length 488;
Best Local Similarity 77.8%; Pred. No. 61;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 STEGFFSGR 9
Db 354 ATEGFTSGR 362
:|||||

RESULT 14
Q8K0F7 PRELIMINARY; PRT; 496 AA.
ID Q8K0F7

AC Q8K0F7;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to tripartite motif protein 39.
GN TRIM39.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast tumor;
RA Strausberg R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR EMBL: BC031540; AAKH1540.1; -.
DR MGD; MGI:1890659; Trim39.
DR InterPro; IPR001870; B302.
DR InterPro; IPR006574; PRY.
DR InterPro; IPR003877; SPRY receptor.
DR InterPro; IPR000315; Znf Bbox.
DR InterPro; IPR001841; Znf_ring.
DR Pfam; PF00622; SPRY, 1.
DR Pfam; PF00643; zf-B_box; 1.
DR Pfam; PF00097; zf-C3HC4; 1.
DR SMART; SM00336; BBOX; 1.
DR SMART; SM00589; PRY; 1.
DR SMART; SM00184; RING; 1.
DR SMART; SM00449; SPRY; 1.
DR PROSITE; PS00119; ZF_BOX; 1.
DR PROSITE; PS00518; ZF_RING_1; 1.
DR PROSITE; PS00089; ZF_RING_2; 1.
KW Metal-binding; Zinc; Zinc-finger.
SQ SEQUENCE 496 AA; 57257 MW; 937B2558186C3D17 CRC64;

Query Match 76.6%; Score 36; DB 11; Length 496;
Best Local Similarity 77.8%; Pred. No. 62;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 STEGFFSGR 9
Db 362 ATEGFTSGR 370
:|||||

RESULT 15
Q8U8L7 PRELIMINARY; PRT; 651 AA.
ID Q8U8L7

Q8U8L7;
AC Q8U8L7;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Glycogen debranching enzyme.
GN GLGX OR ATU4073 OR AGR_L_1566.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Secubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Chena Y.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,
RA Kren V., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D., Str.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58.";
RL Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M., Mullin L.,
RA Gurello B., Goldman B.S., Cao Y., Askenazi O., Iartchouk O., Epp A., Liu F.,
RA Houmel K., Gordon J., Vaudin M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Wollam C., Allinger M., Doughty J., Lomo C., Sear C., Strub G.,
RA Planagan C., Crowell C., Gursen J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58.";
RL Science 294:2323-2328(2001).
DR EMBL; AB009338; AAL44874.1; -.
DR EMBL; AB008277; AAK89356.1; -.
DR InterPro; IPR006047; Alpha_amyl_cat.
DR InterPro; IPR004193; Glyco_hydro_13N.
DR Pfam; PF00128; alpha-amylase; 1.
DR Pfam; PF02922; isoamylase; 1.
KW Complete proteome.
SQ SEQUENCE 651 AA; 71336 MW; 2CBA8D6B398A7B1 CRC64;

Query Match 76.6%; Score 36; DB 16; Length 651;
Best Local Similarity 77.8%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 STEGFFSGR 9
Db 556 SESGFFSGR 564
:|||||

Search completed: December 11, 2003, 18:28:00
Job time : 25.3333 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 11, 2003, 17:22:25 ; Search time 25 Seconds
(without alignments)
57.142 Million cell updates/sec

Title: US-10-014-658-6

Perfect score: 46

Sequence: 1 STREGSGR 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A Geneseq_19Jun03.*

1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	46	100.0	9	21	AA194470 Human antithrombin
2	41	89.1	179	22	AB14761 Human nervous syst
3	41	89.1	677	22	ABG16233 Novel human diagno
4	40	87.0	9	21	AA194469 Human antithrombin
5	40	87.0	9	21	AA194472 Human antithrombin
6	39	84.8	152	21	AA194472 Arabidopsis thalia
7	39	84.8	152	21	AA194472 Arabidopsis thalia
8	39	84.8	163	21	AA194472 Arabidopsis thalia
9	39	84.8	163	21	AA194472 Arabidopsis thalia

10	39	84.8	174	21	AA194470 Arabidopsis thalia
11	39	84.8	174	21	AA194470 Arabidopsis thalia
12	39	84.8	174	23	AA194470 Arabidopsis thalia
13	39	84.8	174	24	AA194470 Arabidopsis thalia
14	38	82.6	1071	15	AA194470 Arabidopsis thalia
15	37	80.4	25	22	AA194470 Arabidopsis thalia
16	36	78.3	914	22	AA194470 Arabidopsis thalia
17	36	78.3	1026	21	AA194470 Arabidopsis thalia
18	36	78.3	1026	22	AA194470 Arabidopsis thalia
19	35	76.1	609	22	AA194470 Arabidopsis thalia
20	35	76.1	1527	20	AA194470 Arabidopsis thalia
21	35	76.1	1527	23	AA194470 Arabidopsis thalia
22	35	76.1	1527	23	AA194470 Arabidopsis thalia
23	35	76.1	1527	23	AA194470 Arabidopsis thalia
24	35	76.1	1528	18	AA194470 Arabidopsis thalia
25	34	73.9	9	21	AA194470 Arabidopsis thalia
26	34	73.9	9	21	AA194470 Arabidopsis thalia
27	34	73.9	57	23	AA194470 Arabidopsis thalia
28	34	73.9	72	22	AA194470 Arabidopsis thalia
29	34	73.9	93	23	AA194470 Arabidopsis thalia
30	34	73.9	117	22	AA194470 Arabidopsis thalia
31	34	73.9	124	22	AA194470 Arabidopsis thalia
32	34	73.9	124	23	AA194470 Arabidopsis thalia
33	34	73.9	127	21	AA194470 Arabidopsis thalia
34	34	73.9	127	22	AA194470 Arabidopsis thalia
35	34	73.9	178	22	AA194470 Arabidopsis thalia
36	34	73.9	178	22	AA194470 Arabidopsis thalia
37	34	73.9	233	22	AA194470 Arabidopsis thalia
38	34	73.9	394	19	AA194470 Arabidopsis thalia
39	34	73.9	402	18	AA194470 Arabidopsis thalia
40	34	73.9	411	22	AA194470 Arabidopsis thalia
41	34	73.9	435	22	AA194470 Arabidopsis thalia
42	34	73.9	441	22	AA194470 Arabidopsis thalia
43	34	73.9	441	22	AA194470 Arabidopsis thalia
44	34	73.9	473	17	AA194470 Arabidopsis thalia
45	34	73.9	473	19	AA194470 Arabidopsis thalia

ALIGNMENTS

RESULT 1
AA194470
ID AA194470 standard; peptide; 9 AA.
XX
AC AA194470;
XX
DT 27-MAR-2000 (first entry)
XX
DE Human antithrombin III variant Bb.B (385-393).
XX
DE Human; antithrombin III; ATIII variant Bb.B; elastase-resistant;
XX Igg activated neutrophil resistant; anti-thrombin activity; heparin;
XX anti-factor Xa activity; blood clotting disorder; sepsis; trauma; stroke;
XX thrombin activation-related pathological symptom; reatenosis; thrombosis;
XX acute respiratory distress syndrome; thromboembolism; reocclusion.
XX
XX Homo sapiens.
OS Synthetic.

Key	Location/Qualifiers
FT Misc-difference 3	/note= "ATIII.N135A Ala at 387 is substituted by Glu"
FT Misc-difference 4	/note= "ATIII.N135A Val at 388 is substituted by Gly"
FT Misc-difference 5	/note= "ATIII.N135A Val at 389 is substituted by Glu"
FT Misc-difference 6	/note= "ATIII.N135A Ile at 390 is substituted by Gly"
FT Misc-difference 7	/note= "ATIII.N135A Ala at 391 is substituted by Ser"
XX	WO9958098-A2.

02-OCT-2000; 2000US-02336802.
 02-OCT-2000; 2000US-0237037.
 02-OCT-2000; 2000US-0237037.
 02-OCT-2000; 2000US-0237039.
 02-OCT-2000; 2000US-0237040.
 02-OCT-2000; 2000US-0239935.
 02-OCT-2000; 2000US-0239937.
 02-OCT-2000; 2000US-0240960.
 02-OCT-2000; 2000US-0241785.
 02-OCT-2000; 2000US-0241786.
 02-OCT-2000; 2000US-0241787.
 02-OCT-2000; 2000US-0241808.
 02-OCT-2000; 2000US-0241809.
 02-OCT-2000; 2000US-0241826.
 02-OCT-2000; 2000US-0242221.
 01-NOV-2000; 2000US-0246474.
 08-NOV-2000; 2000US-0246475.
 08-NOV-2000; 2000US-0246476.
 08-NOV-2000; 2000US-0246477.
 08-NOV-2000; 2000US-0246528.
 08-NOV-2000; 2000US-0246532.
 08-NOV-2000; 2000US-0246533.
 08-NOV-2000; 2000US-0246523.
 08-NOV-2000; 2000US-0246524.
 08-NOV-2000; 2000US-0246525.
 08-NOV-2000; 2000US-0246526.
 08-NOV-2000; 2000US-0246527.
 08-NOV-2000; 2000US-0246528.
 08-NOV-2000; 2000US-0246532.
 08-NOV-2000; 2000US-0246533.
 08-NOV-2000; 2000US-0246610.
 08-NOV-2000; 2000US-0246611.
 08-NOV-2000; 2000US-0246613.
 08-NOV-2000; 2000US-0246617.
 17-NOV-2000; 2000US-0249207.
 17-NOV-2000; 2000US-0249208.
 17-NOV-2000; 2000US-0249209.
 17-NOV-2000; 2000US-0249210.
 17-NOV-2000; 2000US-0249211.
 17-NOV-2000; 2000US-0249212.
 17-NOV-2000; 2000US-0249213.
 17-NOV-2000; 2000US-0249214.
 17-NOV-2000; 2000US-0249215.
 17-NOV-2000; 2000US-0249216.
 17-NOV-2000; 2000US-0249217.
 17-NOV-2000; 2000US-0249218.
 17-NOV-2000; 2000US-0249244.
 17-NOV-2000; 2000US-0249245.
 17-NOV-2000; 2000US-0249246.
 17-NOV-2000; 2000US-0249285.
 17-NOV-2000; 2000US-0249297.
 17-NOV-2000; 2000US-0249299.
 17-NOV-2000; 2000US-0249300.
 01-DEC-2000; 2000US-0250391.
 01-DEC-2000; 2000US-0251160.
 05-DEC-2000; 2000US-0251030.
 05-DEC-2000; 2000US-0251988.
 05-DEC-2000; 2000US-0256719.
 06-DEC-2000; 2000US-0251479.
 08-DEC-2000; 2000US-0251856.
 08-DEC-2000; 2000US-0251868.
 08-DEC-2000; 2000US-0251869.
 08-DEC-2000; 2000US-0251989.
 08-DEC-2000; 2000US-0251990.
 11-DEC-2000; 2000US-0254097.
 05-JAN-2001; 2001US-0259678.
 (HUMA-) HUMAN GENOME SCI INC.
 Rosen CA, Baraah SC, Ruben SM;
 WPI; 2001-541565/60.
 N-PSDB; ABA11087.
 Nucleic acids encoding 3224 human nervous system antigen polypeptides,
 02-OCT-2000; 2000US-02336802.
 02-OCT-2000; 2000US-0237037.
 02-OCT-2000; 2000US-0237037.
 02-OCT-2000; 2000US-0237039.
 02-OCT-2000; 2000US-0237040.
 02-OCT-2000; 2000US-0239935.
 02-OCT-2000; 2000US-0239937.
 02-OCT-2000; 2000US-0240960.
 02-OCT-2000; 2000US-0241785.
 02-OCT-2000; 2000US-0241786.
 02-OCT-2000; 2000US-0241787.
 02-OCT-2000; 2000US-0241808.
 02-OCT-2000; 2000US-0241809.
 02-OCT-2000; 2000US-0241826.
 02-OCT-2000; 2000US-0242221.
 01-NOV-2000; 2000US-0246474.
 08-NOV-2000; 2000US-0246475.
 08-NOV-2000; 2000US-0246476.
 08-NOV-2000; 2000US-0246477.
 08-NOV-2000; 2000US-0246528.
 08-NOV-2000; 2000US-0246532.
 08-NOV-2000; 2000US-0246533.
 08-NOV-2000; 2000US-0246523.
 08-NOV-2000; 2000US-0246524.
 08-NOV-2000; 2000US-0246525.
 08-NOV-2000; 2000US-0246526.
 08-NOV-2000; 2000US-0246527.
 08-NOV-2000; 2000US-0246528.
 08-NOV-2000; 2000US-0246532.
 08-NOV-2000; 2000US-0246533.
 08-NOV-2000; 2000US-0246610.
 08-NOV-2000; 2000US-0246611.
 08-NOV-2000; 2000US-0246613.
 08-NOV-2000; 2000US-0246617.
 17-NOV-2000; 2000US-0249207.
 17-NOV-2000; 2000US-0249208.
 17-NOV-2000; 2000US-0249209.
 17-NOV-2000; 2000US-0249210.
 17-NOV-2000; 2000US-0249211.
 17-NOV-2000; 2000US-0249212.
 17-NOV-2000; 2000US-0249213.
 17-NOV-2000; 2000US-0249214.
 17-NOV-2000; 2000US-0249215.
 17-NOV-2000; 2000US-0249216.
 17-NOV-2000; 2000US-0249217.
 17-NOV-2000; 2000US-0249218.
 17-NOV-2000; 2000US-0249244.
 17-NOV-2000; 2000US-0249245.
 17-NOV-2000; 2000US-0249246.
 17-NOV-2000; 2000US-0249285.
 17-NOV-2000; 2000US-0249297.
 17-NOV-2000; 2000US-0249299.
 17-NOV-2000; 2000US-0249300.
 01-DEC-2000; 2000US-0250391.
 01-DEC-2000; 2000US-0251160.
 05-DEC-2000; 2000US-0251030.
 05-DEC-2000; 2000US-0251988.
 05-DEC-2000; 2000US-0256719.
 06-DEC-2000; 2000US-0251479.
 08-DEC-2000; 2000US-0251856.
 08-DEC-2000; 2000US-0251868.
 08-DEC-2000; 2000US-0251869.
 08-DEC-2000; 2000US-0251989.
 08-DEC-2000; 2000US-0251990.
 11-DEC-2000; 2000US-0254097.
 05-JAN-2001; 2001US-0259678.
 (HUMA-) HUMAN GENOME SCI INC.
 Rosen CA, Baraah SC, Ruben SM;
 WPI; 2001-541565/60.
 N-PSDB; ABA11087.
 Nucleic acids encoding 3224 human nervous system antigen polypeptides,
 useful for preventing, diagnosing and/or treating nervous system
 cancers and metastases -
 Claim 11; SEQ ID NO 3418; 1701pp + Sequence Listing; English.
 The invention relates to novel genes (ABA11004-ABA21534) and proteins
 (ABA14678-ABA18001) useful for preventing, treating or ameliorating
 medical conditions e.g. by protein or gene therapy. The genes are
 isolated from a range of human tissues disclosed in the specification.
 The nucleic acids, proteins, antibodies and (ant)agonists are useful
 in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 and ovarian cancer and other cancers of the adrenal gland, bone, bone
 marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 colitis; (c) cardiovascular disorders such as myocardial ischaemias;
 (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 and parasitic infections.
 Note: The sequence data for this patent did not form part of the
 printed specification, but was obtained in electronic format directly
 from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 Query Match 89.1%; Score 41; DB 22; Length 179;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 STEGEGSG 8
 |||||
 DB 58 STEGEGSG 65
 RESULT 3
 ABG16233
 ID ABG16233 standard; Protein; 677 AA.
 XX
 AC ABG16233;
 XX
 DT 18-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #16224.
 XX
 KW Human; chromosome mapping; gene mapping; Gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR N-PSDB; AAS80420.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 20; SEQ ID No 46592; 103pp; English.
 XX

PN W09958098-A2.
 XX 18-NOV-1999.
 XX 12-MAY-1999; 99WO-US10549.
 XX 12-MAY-1998; 99US-0085197.
 XX 05-MAY-1999; 99US-0085197.
 XX (BOCK/) BOCK S C.
 XX (BICA/) PICARD V.
 XX (ZEND/) ZENDEHROUH P.
 PI Bock SC, Picard V, Zendeihrouh P;
 XX WPI; 2000-116274/10.
 XX
 XX New modified human antithrombin III compounds, used for treating e.g.
 PT sepsis, trauma, acute respiratory distress syndrome, restenosis,
 PT thrombosis, thromboembolism or stroke -
 XX
 XX Claim 13; Page 57; 75pp; English.
 XX
 XX The present sequence is from an antithrombin III (ATIII) variant, Bb.D
 CC derived from human ATIII N135A cDNA insert of the pBluebac baculovirus
 CC expression construct and comprises residues 385-393. The variant has
 CC improved resistance to elastase and IgG-activated neutrophils while
 CC retains anti-thrombin and anti-factor Xa activities. It may be
 CC expressed as glycoforms with enhanced heparin affinity which target the
 CC blood vessel wall more efficiently than ATIIIs with normal heparin
 CC affinity. The modified ATIIIs can be used to treat thrombin activation-
 CC related pathological symptoms due to sepsis, trauma, acute
 CC respiratory distress syndrome, restenosis, thrombosis, thromboembolism
 CC and stroke. It can also be used to reduce the risk of reocclusion
 CC and restenosis in percutaneous transluminal coronary angioplasty,
 CC thrombosis associated with surgery, ischaemia/reperfusion injury, and
 CC coagulation abnormalities in cancer or surgical patients.
 XX
 XX Sequence 9 AA;
 SQ
 Query Match 87.08; Score 40; DB 21; Length 9;
 Best Local Similarity 88.99; Pred. NO. 9.3e+05;
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 STEGEGSGR 9
 Db 1 SEEGEGSGR 9
 RESULT 6
 AAG11548
 ID AAG11548 standard; Protein; 152 AA.
 XX
 XX AAG11548;
 XX
 XX 17-OCT-2000 (first entry)
 XX
 XX Arabidopsis thaliana protein fragment SEQ ID NO: 10305.
 XX
 XX Protein identification; signal transduction pathway; metabolic pathway;
 XX hybridisation assay; genetic mapping; gene expression control; promoter;
 XX termination sequence.
 XX
 XX Arabidopsis thaliana.
 OS
 XX EP1033405-A2.
 XX
 XX 06-SEP-2000.
 PD
 XX 25-FEB-2000; 2000EP-0301439.
 XX
 XX 25-FEB-1999; 99US-0121825.
 XX
 XX 05-MAR-1999; 99US-0123180.
 PR
 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127482.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 23-APR-1999; 99US-0130891.
 PR 28-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 06-MAY-1999; 99US-0132487.
 PR 07-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 24-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.

PR	16-APR-1999;	99US-0129845.
PR	19-APR-1999;	99US-0130077.
PR	21-APR-1999;	99US-0130449.
PR	23-APR-1999;	99US-0130510.
PR	23-APR-1999;	99US-0130891.
PR	28-APR-1999;	99US-0131449.
PR	30-APR-1999;	99US-0132048.
PR	30-APR-1999;	99US-0132407.
PR	04-MAY-1999;	99US-0132484.
PR	05-MAY-1999;	99US-0132485.
PR	06-MAY-1999;	99US-0132486.
PR	06-MAY-1999;	99US-0132487.
PR	07-MAY-1999;	99US-0132863.
PR	11-MAY-1999;	99US-0134256.
PR	14-MAY-1999;	99US-0134218.
PR	14-MAY-1999;	99US-0134219.
PR	14-MAY-1999;	99US-0134221.
PR	14-MAY-1999;	99US-0134370.
PR	18-MAY-1999;	99US-0134768.
PR	19-MAY-1999;	99US-0134941.
PR	20-MAY-1999;	99US-0135124.
PR	21-MAY-1999;	99US-0135353.
PR	24-MAY-1999;	99US-0135629.
PR	25-MAY-1999;	99US-0136021.
PR	27-MAY-1999;	99US-0136392.
PR	28-MAY-1999;	99US-0136782.
PR	01-JUN-1999;	99US-0137222.
PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138847.
PR	14-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	16-JUN-1999;	99US-0139453.
PR	17-JUN-1999;	99US-0139492.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	18-JUN-1999;	99US-0139763.
PR	21-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	23-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140695.
PR	28-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140931.
PR	30-JUN-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142154.
PR	02-JUL-1999;	99US-0142055.
PR	06-JUL-1999;	99US-0142390.
PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	19-JUL-1999;	99US-0144322.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0129845.
PR	19-JUL-1999;	99US-0130077.
PR	20-JUL-1999;	99US-0130449.
PR	20-JUL-1999;	99US-0130510.
PR	20-JUL-1999;	99US-0130891.
PR	21-JUL-1999;	99US-0131449.
PR	21-JUL-1999;	99US-0132048.
PR	21-JUL-1999;	99US-0132407.
PR	22-JUL-1999;	99US-0132484.
PR	22-JUL-1999;	99US-0132485.
PR	22-JUL-1999;	99US-0132486.
PR	22-JUL-1999;	99US-0132487.
PR	23-JUL-1999;	99US-0132863.
PR	23-JUL-1999;	99US-0134256.
PR	23-JUL-1999;	99US-0134218.
PR	26-JUL-1999;	99US-0134219.
PR	27-JUL-1999;	99US-0134221.
PR	27-JUL-1999;	99US-0134370.
PR	27-JUL-1999;	99US-0134768.
PR	28-JUL-1999;	99US-0134941.
PR	02-AUG-1999;	99US-0135124.
PR	02-AUG-1999;	99US-0135353.
PR	02-AUG-1999;	99US-0135629.
PR	03-AUG-1999;	99US-0136021.
PR	04-AUG-1999;	99US-0136392.
PR	04-AUG-1999;	99US-0136782.
PR	05-AUG-1999;	99US-0137222.
PR	06-AUG-1999;	99US-0137528.
PR	06-AUG-1999;	99US-0137502.
PR	09-AUG-1999;	99US-0137724.
PR	09-AUG-1999;	99US-0138094.
PR	10-AUG-1999;	99US-0138540.
PR	10-AUG-1999;	99US-0138847.
PR	11-AUG-1999;	99US-0139119.
PR	12-AUG-1999;	99US-0139452.
PR	13-AUG-1999;	99US-0139453.
PR	13-AUG-1999;	99US-0139492.
PR	16-AUG-1999;	99US-0139454.
PR	17-AUG-1999;	99US-0139455.
PR	18-AUG-1999;	99US-0139456.
PR	20-AUG-1999;	99US-0139457.
PR	20-AUG-1999;	99US-0139458.
PR	20-AUG-1999;	99US-0139459.
PR	23-AUG-1999;	99US-0139460.
PR	23-AUG-1999;	99US-0139461.
PR	26-AUG-1999;	99US-0139462.
PR	26-AUG-1999;	99US-0150566.
PR	27-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	13-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	23-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0158293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.